Abstract Algebra

Romyar Sharifi

Contents

Introduction

In mathematics, we often encounter objects that are sets with various operations that can be performed on them. For instance, one may add and multiply integers, and one can do the same with rational numbers, real numbers, and even complex (or imaginary) numbers. Or, given two functions that input and output real numbers, we can compose them. We can add vectors, or multiply them by scalars.

In abstract algebra, we attempt to provide lists of properties that common mathematical objects satisfy. Given such a list of properties, we impose them as "axioms", and we study the properties of objects that satisfy these axioms. The objects that we deal with most in the first part of these notes are called groups, rings, and fields.

Groups, rings, and fields all sets with binary operations. A binary operation inputs two elements of the set and outputs a third such element. Addition and multiplication of integers, for instance, are binary operations, as is composition of real-valued functions of a real number. Scalar multiplication of a vector in the plane is not however, since it starts not with two vectors, but rather a scalar (i.e., a real number) and a vector.

We often require our binary operations to have certain properties like associativity or commutativity. If we call our operation "$\star$", then associativity reads

$$(x \star y) \star z = x \star (y \star z),$$

and commutativity reads

$$x \star y = y \star x.$$

In imprecise terms, they tell you that the order in which you perform the operations doesn't matter. Though the most typically-encountered binary operations tend to be associative, many are not commutative (i.e., are "noncommutative"). For instance, you may recall that the order of composition of functions matters: e.g., $\sin(x^2)$ and $\sin^2(x)$ are two different things.

The integers $\mathbb{Z}$, the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$ are all rings, as is the set of n by n matrices with entries in any of these. A ring is a set with two binary operations called addition and multiplication. In order to be a ring, we require associativity of both operations, commutativity of addition, and distributivity of the two operations. Distributivity can be expressed as follows:

$$(x+y) \cdot (z+w) = x \cdot z + x \cdot w + y \cdot z + y \cdot w.$$

Every ring must have an element called "0" (satisfying $0+x=x$ for any x) and, for every number x, there should be another $-x$ which when added to x, gives you 0. Typically, a ring also has an element called "1", which satisfies $1 \cdot x = x = x \cdot 1$. For instance, in any ring of n by n matrices, the element "1" is actually the identity matrix.

You may recall that multiplication of square matrices is noncommutative (if they are at least 2 by 2 in size). For instance, we have

$$\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}\cdot\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}=\begin{pmatrix}1 & -1\\ 0 & -1\end{pmatrix}\neq\begin{pmatrix}1 & 1\\ 0 & -1\end{pmatrix}=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\cdot\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}.$$

On the other hand, the multiplications in $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all commutative. A ring in which multiplication is a commutative binary operation is a called a commutative ring.

Once we have rings, fields are simple to describe. Fields are commutative rings with one extra property. That is, a field has inverses under multiplication: if x is in the field and isn't 0, then there must be an element $x^{-1} = 1/x$ as well, and it satisfies $x \cdot x^{-1} = 1$. In particular, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields as well as rings, but $\mathbb{Z}$ is not a field. In a field, fractions add and multiply in the familiar way:

$$\frac{x}{y}+\frac{z}{w}=\frac{xw+yz}{yw} \quad\text{and}\quad \frac{x}{y}\cdot\frac{z}{w}=\frac{xz}{yw}.$$

Some rings have nonzero elements x and y with product xy equal to 0. These are called zero-divisors. For instance,

$$\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\cdot\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}=\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix},$$

and so we can have that the product of two nonzero matrices is the zero matrix. If a commutative ring has no zero divisors, then we can construct its field of fractions artificially. Its elements consisting of elements denoted x/y, where x and y are in the original rings. The field of fractions of $\mathbb{Z}$ is $\mathbb{Q}$, and here we have our first example of a construction that is well-known for the simplest ring of all, the integers, but can be performed more generally (for instance to polynomials), starting from the axioms of a ring and a few extra properties.

Groups may seem a bit less familiar, but they are also in a sense simpler. Groups have only one binary operation. Call it whatever you like: addition, multiplication, or just "$\star$". A group and its binary operation $\star$ must satisfy just three properties: associativity of $\star$, the existence of an identity element e, and the existence of inverses. The identity element e is like the number 1 is under multiplication, or like 0 is under addition, in the rings that are familiar to us. It satisfies

$$e \star x = x = x \star e$$

for all x in the group. The inverse of an element x is normally denoted x^{-1}, but it is written $-x$ if our operation is addition. It satisfies

$$x^{-1} \star x = e = x \star x^{-1}.$$

In particular, rings are groups if we forget about the multiplication and just consider the operation of addition. Fields are groups under multiplication if we throw out 0.

Many less familiar but interesting mathematical objects are groups. The rotations of a circle form a group under composition (following one rotation by another), and the permutations (switches of positions) of five balls between five slots are a group under composition as well. The n by n real matrices with nonzero determinant form a group under multiplication too. The set of "moves" of a Rubik's cube (compositions of rotations of sides by 90 degree multiples) form a group too: a very complicated one, in fact. So, groups are in some sense a less refined but much broader class of objects than the rings, with more exotic members.

In our examples, some of the groups have finitely many elements and hence are known as finite groups. Here's an interesting property of every finite group. Suppose that a finite group G has n elements, and let x be one of them. Then x^n, which is $x \star x \star \cdots \star x$ with x appearing n times, is the identity element e. For instance, if I permute the position of 5 balls in five slots in a certain manner, over and over, the balls will wind up in the position they started after 120 steps, since that is the order of the group. In fact, this exaggerates the number of repetitions needed: the balls end up at the starting point in six or fewer. The same goes with the Rubik's cube: repeat the same sequence of moves enough times, and, if you have enough patience (meaning watch out for carpal tunnel syndrome), you will end back up where you started. This is something that, a priori, may not seem obvious at all. Yet, this property of finite groups is a very general phenomenon, derived solely from the group axioms.

Hopefully this encourages you to believe that abstract algebra may be of serious use both inside and outside mathematics, and indeed, it is so, in addition to being a fascinating and beautiful theory in its own right for those so inclined. In the next chapter, we begin our study of abstract algebra at a much more leisurely pace.

Part 1

A First Course

CHAPTER 1

Set theory

1.1. Sets and functions

In these notes, we assume some basic notions from set theory, for which we give only the briefest of reviews. We won't attempt to define a set formally here. Instead, we simply make some remarks about them. Vaguely, a set is a collection of objects. Not every collection of objects is a set: the "collection" of all sets is not a set. On the other hand, most reasonable collections of objects are sets: the integers, the real numbers, the movies in your DVD collection (seemingly, a soon-to-be dated notion), those are sets.

Sets consist of elements. If X is a set, we write $x \in X$ to mean that x is an element of X (or "x is in X"). Similarly, $x \notin X$ means that x is not an element of X (which only really makes sense if both x and the elements of X are elements of some common larger set so they can be compared.)

EXAMPLES 1.1.1.

a. The empty set $\varnothing$ is the set with no elements.

b. The set consisting of elements called 1, 2, and 3 is denoted $\{1,2,3\}$, and this notation extends to any finite collection of objects.

c. The set $\{1,2,3,\dots\}$ of positive integers is again a set.

d. The real numbers $\mathbb{R}$ form a set.

Any collection of elements of a set X is called a subset of X and is a set itself. We write $Y \subseteq X$ to mean that Y is a subset of X. If Y and Z are different subsets of X, then we write $Y \neq Z$ and we say that Y and Z are distinct subsets.

A property P that only some elements of X satisfy allows us to specify a subset of X consisting of elements of X that satisfy P, which we denote in set-theoretic notation by

$$\{x \in X \mid x \text{ satisfies } P\},$$

or just $\{x \mid x \text{ satisfies } P\}$ if X is understood.

EXAMPLE 1.1.2. The subset $\{n \in \mathbb{Z} \mid 2 \text{ divides } n\}$ of $\mathbb{Z}$ is the set of even integers.

DEFINITION 1.1.3. Let X be a set and Y be a subset of X. Then $X - Y$ denotes the *complement* of Y in X, which is defined as

$$X - Y = \{x \in X \mid x \notin Y\}.$$

If Y is a subset of X that is not X itself, then it is called a proper subset, and we write $Y \subset X$ (or $Y \subsetneq X$). Given two subsets Y and Z of a larger set X, we can form their union $Y \cup Z$ and their intersection $Y \cap Z$, which are also subsets of X.

DEFINITION 1.1.4. Given sets X and Y, the *direct product* $X \times Y$ is the set of pairs (x,y) with $x \in X$ and $y \in Y$.

In set-theoretic notation, we may write this as

$$X \times Y = \{(x,y) \mid x \in X, y \in Y\}.$$

DEFINITION 1.1.5. A *function* $f\colon X \to Y$ from a set X to a set Y is a rule that to each $x \in X$ associates a single element $f(x) \in Y$, known as the value of f at x.

NOTATION 1.1.6. We sometimes refer to a function as a *map*, and we sometimes write $f\colon x \mapsto y$ to indicate that $f(x) = y$, or in other words that f *maps* (or *sends*) x to y.

We can, of course, compose functions, as in the following definition.

DEFINITION 1.1.7. Let X, Y, and Z be sets and $f\colon X \to Y$ and $g\colon Y \to Z$ functions. The *composition* (or *composite function*) $g \circ f\colon X \to Z$ of g with f is the function such that $(g \circ f)(x) = g(f(x))$ for all $x \in X$.

DEFINITION 1.1.8. Let $f\colon X \to Y$ be a function.

a. The function f is *one-to-one* (or *injective*, or an *injection*) if for every $x, y \in X$ such that $f(x) = f(y)$, one has $x = y$.

b. The function f is *onto* (or *surjective*, or a *surjection*) if for every $y \in Y$, there exists an $x \in X$ such that $f(x) = y$.

c. The function f a *one-to-one correspondence* (or *bijective*, or a *bijection*) if it is both one-to-one and onto.

REMARK 1.1.9. In other words, to say a function $f\colon X \to Y$ is one-to-one is to say that it sends at most one element of X to any given element of Y. To say that it is onto is to say that it sends at least one element of X to any given element of Y. So, of course, to say that it is a one-to-one correspondence is to say that it sends exactly one element of X to each element of Y.

EXAMPLES 1.1.10.

a. The map $f\colon \mathbb{Z} \to \mathbb{Z}$ defined by $f(x) = 2x$ for every $x \in \mathbb{Z}$ is one-to-one, but not onto.

b. The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ is a bijection.

c. The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x\sin(x)$ is onto, but not one-to-one.

DEFINITION 1.1.11.

a. A set X is *finite* if X has only a finite number of elements, and it is *infinite* otherwise.

b. If X is a finite set, then the *order* $|X|$ of X is the number of elements it has.

PROPOSITION 1.1.12. *Let X and Y be finite sets of the same order, and let $f\colon X \to Y$ be a function. Then f is one-to-one if and only if it is onto.*

PROOF. Let $n = |X|$, and denote the elements of X by $x_1, \ldots, x_n$. If $f(x_i) = f(x_j)$ for some $i \neq j$, then the subset $\{f(x_1), \ldots, f(x_n)\}$ of Y has fewer than n elements, hence cannot equal Y. Conversely, if $\{f(x_1), \ldots, f(x_n)\}$ has fewer than n elements, then $f(x_i) = f(x_j)$ for some $i \neq j$. Therefore f is not one-to-one if and only if it is not onto, as desired. □

Note that every bijection has an inverse.

DEFINITION 1.1.13. If $f\colon X \to Y$ is a bijection, then we define the *inverse* of f to be the function $f^{-1}\colon Y \to X$ satsifying $f^{-1}(y) = x$ for the unique x such that $f(x) = y$.

Given a bijection $f\colon X \to Y$, note that

$$f^{-1}(f(x)) = x \quad \text{and} \quad f(f^{-1}(y)) = y$$

for all $x \in X$ and $y \in Y$. In other words, $f^{-1} \circ f$ (resp., $f \circ f^{-1}$) is the function that takes every element of Y (resp., X) to itself.

EXAMPLE 1.1.14. The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ has inverse $f^{-1}(x) = x^{-1/3}$.

Often, it is useful to use what is called an indexing set I to define a collection, which is just some given set, like the natural numbers. Given objects x_i for each $i \in I$, we can use set-theoretic notation to define a set consisting of them

$$\{x_i \mid i \in I\}$$

that is in one-to-one correspondence with I via the map that takes i to x_i.

DEFINITION 1.1.15. Let X be a set and $\{Y_i \mid i \in I\}$ be a collection of subsets of X indexed by a set I.

a. The *intersection* and *union* of the sets Y_i are defined as

$$\bigcap_{i\in I} Y_i = \{x \in X \mid x \in Y_i \text{ for all } i \in I\} \quad \text{and} \quad \bigcup_{i\in I} Y_i = \{x \in X \mid x \in Y_i \text{ for some } i \in I\},$$

respectively.

b. If $Y_i \cap Y_j = \varnothing$ for every $i, j \in I$ with $i \neq j$, we say that the sets Y_i are *disjoint*.

c. If the collection of Y_i is disjoint, then their union is called a *disjoint union* and is often written as

$$\coprod_{i\in I} Y_i.$$

DEFINITION 1.1.16. Let $\{X_i \mid i \in I\}$ be a collection of sets. The *direct product* $\prod_{i\in I} X_i$ of the X_i is the set of tuples

$$\prod_{i\in I} X_i = \{(x_i)_{i\in I} \mid x_i \in X_i\}.$$

In other words, an element of $\prod_{i\in I} X_i$ is a choice of one element of X_i for each $i \in I$.

1.2. Relations

In this section, we consider several types of a very general construct called a relation.

DEFINITION 1.2.1. A *relation* R is a subset of $X \times Y$. We often write xRy to indicate that $(x, y) \in R$.

EXAMPLES 1.2.2.

a. The circle $S^1 = \{x^2 + y^2 = 1\}$ forms a relation in $\mathbb{R} \times \mathbb{R}$. As is well-known, a pair (x,y) is in S^1 if and only if $(x,y) = (\cos\theta, \sin\theta)$ for some $\theta \in [0, 2\pi)$.

b. The relation $\leq$ on $\mathbb{R} \times \mathbb{R}$ is given by $\{(x,y) \mid x \leq y\}$.

As a first example, we see that functions can be considered as relations.

REMARK 1.2.3. A function $f \colon X \to Y$ gives rise to a relation

$$\Gamma_f = \{(x, f(x)) \mid x \in X\} \subseteq X \times Y,$$

known as the graph of f. Equivalently, each relation R in $X \times Y$ such that for each $x \in X$ there exists a unique $y \in Y$ with xRy gives rise to a function f defined by $f(x) = y$ (where xRy).

EXAMPLE 1.2.4. The relation in $\mathbb{R}^2$ corresponding to $f \colon \mathbb{R} \to \mathbb{R}$ is the graph of f in the usual sense.

We will consider two other types of relations.

DEFINITION 1.2.5. An *equivalence relation* $\sim$ on X is a relation in $X \times X$ that satisfies the following properties.

a. (*reflexivity*) For all $x \in X$, we have $x \sim x$.

b. (*symmetry*) For any $x, y \in X$, we have $x \sim y$ if and only if $y \sim x$.

c. (*transitivity*) If $x, y, z \in X$ satisfy $x \sim y$ and $y \sim z$, then $x \sim z$.

EXAMPLES 1.2.6.

a. Equality is an equivalence relation $=$ on any set X. As a relation, it defines the subset $\{(x,x) \mid x \in X\}$ of $X \times X$.

b. The relation $\leq$ on $\mathbb{R}$ is not an equivalence relation, as it is not symmetric.

c. Let n be a positive integer, and consider the relation $\equiv_n$ on $\mathbb{Z}$ defined by $a \equiv_n b$ if $a - b$ is divisible by n. This is an equivalence relation known as congruence modulo n. We will write $a \equiv b \bmod n$ in place of $a \equiv_n b$, as is standard.

DEFINITION 1.2.7. Let $\sim$ be an equivalence relation on a set X. The *equivalence class* of $x \in X$ is the set $\{y \in X \mid x \sim y\}$.

EXAMPLES 1.2.8.

a. The equivalence classes under $=$ on a set X are just the singleton sets $\{x\}$ for $x \in X$.

b. The equivalence class of 3 under $\equiv_7$ on $\mathbb{Z}$ is $\{\ldots, -11, -4, 3, 10, 17, \ldots\}$.

DEFINITION 1.2.9. We refer to the set of equivalence classes of $\mathbb{Z}$ under congruence modulo n as the *integers modulo n*, and denote it $\mathbb{Z}/n\mathbb{Z}$. (Note that number theorists usually denote this set $\mathbb{Z}/n\mathbb{Z}$.) A typical member $\overline{a}$ has the form

$$\overline{a} = \{a + bn \mid b \in \mathbb{Z}\}$$

for some integer a. An equivalence class $\overline{a}$ is known as a congruence class modulo n.

LEMMA 1.2.10. *The distinct equivalence classes of X under an equivalence relation $\sim$ are disjoint, and X is the disjoint union of its distinct equivalence classes.*

PROOF. The second statement follows from the first once we known that different equivalence classes are disjoint, since every $x \in X$ is in some equivalence class. For the first statement, suppose that x and y are elements of X, and let E_x and E_y denote their respective equivalence classes under $\sim$. If E_x and E_y are not disjoint, then there exists $z \in E_x \cap E_y$, so $x \sim z$ and $y \sim z$. But then $z \sim x$ by symmetry of $\sim$, and so for any $w \in X$, we have $x \sim w$ implies $z \sim w$ by transitivity of $\sim$. Given that and using $y \sim z$, we then have $y \sim w$, again by transitivity. Hence $E_x \subseteq E_y$. But since x and y are interchangeable in the last sentence, we have $E_y \subseteq E_x$ as well. Therefore, $E_x = E_y$, which is to say any two equivalence classes of X are either distinct or equal. □

DEFINITION 1.2.11. Let X be a set and $\sim$ an equivalence relation on X.

a. For any equivalence class E of $\sim$, a *representative* of E is just an element of E.

b. A *set of representatives* (of the equivalence classes) of X under $\star$ is a subset S of X such that each equivalence class of X contains exactly one element of S.

EXAMPLE 1.2.12. The set $\{0,1,2,\dots,n-1\}$ is a set of representatives of $\mathbb{Z}$ under congruence modulo n.

DEFINITION 1.2.13. A *partial ordering* on a set X is a relation $\leq$ in $X \times X$ that satisfies the following properties.

i. (*reflexivity*) For all $x \in X$, we have $x \leq x$.

ii. (*antisymmetry*) If $x,y \in X$ satisfy $x \leq y$ and $y \leq x$, then $x = y$.

iii. (*transitivity*) If $x,y,z \in X$ satisfy $x \leq y$ and $y \leq z$, then $x \leq z$.

A set X together with a partial ordering $\leq$ is referred to as a *partially ordered set*.

EXAMPLES 1.2.14.

a. The relation $\leq$ on $\mathbb{R}$ is a partial ordering, as is $\geq$.

b. The relation $<$ on $\mathbb{R}$ is not a partial ordering, as it is not reflexive.

c. The relation $\subseteq$ on the set of subsets $\mathscr{P}_X$ of any set X, which is known as the *power set* of X, is a partial ordering.

d. The relation $=$ is a partial ordering on any set.

e. The relation $\equiv_n$ is not a partial ordering on $\mathbb{Z}$, as 0 and n are congruent, but not equal.

Given a partial ordering $\leq$ on a set X, we can speak of minimal and maximal elements of X.

DEFINITION 1.2.15. Let X be a set with a partial ordering $\leq$.

a. A *minimal element* in X (under $\leq$) is an element $x \in X$ such that if $z \in X$ and $z \leq x$, then $z = x$.

b. A *maximal element* $y \in X$ is an element such that if $z \in X$ and $y \leq z$, then $z = y$.

Minimal and maximal elements need not exist, and when they exist, they need not be unique. Here are some examples.

EXAMPLES 1.2.16.

a. The set $\mathbb{R}$ has no minimal or maximal elements under $\leq$.

b. The interval $[0,1)$ in $\mathbb{R}$ has the minimal element 0 but no maximal element under $\leq$.

c. The power set $\mathscr{P}_X$ of X has the minimal element $\varnothing$ and maximal element X under $\subseteq$.

d. Under $=$ on X, every element is both minimal and maximal.

e. Consider the set S of nonempty sets of the integers $\mathbb{Z}$, with partial ordering $\subseteq$. The minimal elements of S are exactly the singleton sets $\{n\}$ for $n \in \mathbb{Z}$.

One can ask for a condition under which maximal (or minimal) elements exist. To phrase such a condition, we need two more notions.

DEFINITION 1.2.17. Let X be a set with a partial ordering $\leq$. A *chain* in X is a subset C of X such that if $x, y \in C$, then either $x \leq y$ or $y \leq x$.

That is, a chain is a subset under which every two elements can be compared by the partial ordering.

EXAMPLE 1.2.18. The power set $\mathscr{P}_X$ of $X = \{1,2,3\}$ is not a chain, as we have neither $\{1,2\}$ contained in $\{2,3\}$, nor $\{2,3\}$ contained in $\{1,2\}$. However, its subset $\{\varnothing, \{1\}, \{1,2\}, \{1,2,3\}\}$ does form a chain.

EXAMPLE 1.2.19. Any subset of $\mathbb{R}$ forms a chain under $\leq$.

The previous example leads us to the following definition, which we mention primarily as a remark.

DEFINITION 1.2.20. If X is itself a chain under $\leq$, then $\leq$ is said to be a *total ordering* on X.

We need the notion of bounds on subsets of a partially ordered set.

DEFINITION 1.2.21. Let X be a set with a partial ordering $\leq$. Let A be a subset of X. An *upper bound* on A under $\leq$ is an element $x \in X$ such that $a \leq x$ for all $a \in A$.

That is, an upper bound on a subset is an element of the set that is at least as large as every element in the subset. Note that the upper bound need not, but can, be contained in the subset itself. (And, of course, lower bounds could have been defined similarly.)

EXAMPLES 1.2.22.

a. The subset $[0,1)$ of $\mathbb{R}$ has an upper bound $1 \in \mathbb{R}$ under $\leq$. In fact, any element $x \geq 1$ is an upper bound for $[0,1)$. The subset $[0,1]$ has the same upper bounds.

b. The subset $\mathbb{Q}$ of $\mathbb{R}$ has no upper bound under $\leq$.

We now come to Zorn's lemma, which is equivalent to the so-called "axiom of choice", and as such, is as much an axiom as it is a theorem (and more of a theorem than it is a lemma). Some, though far from most, mathematicians prefer not to include the axiom of choice among the axioms of set theory, fearing that the resulting collection of axioms may be logically incompatible. For the purposes of this book, we will have no such qualms, and we state Zorn's lemma without proof: the reader may take it as an axiom.

THEOREM 1.2.23 (Zorn's lemma). *Let X be a set with a partial ordering $\leq$, and suppose that every chain in X has an upper bound. Then X contains a maximal element.*

Later on in these notes, we will see a couple of examples where Zorn's lemma can be used to produce the existence of maximal elements in situations of use to algebraists. Zorn's lemma is the form of the axiom of choice considered most conducive to applications in algebra.

Finally, let us consider the notion of generation. We have the following rather obvious lemma.

LEMMA 1.2.24. *Let X be a set and S be a subset. Let P be a subset of $\mathscr{P}_X$ containing X such that P is closed under intersection, and let P_S be the (nonempty) subset of elements of P containing S. Then the intersection of the elements of P_S is the unique minimal element of P_S. That is, it is the smallest subset of X in P containing S.*

We think of P of some property of certain subsets of X that X itself satisfies, where a subset of X is in P it has the property. As P is closed under intersection, for any subset S of X, we may speak of the smallest subset that contains S and has property P. We then think of this subset as the subset of X with property P generated by S. For instance, we have the following.

EXAMPLE 1.2.25. Let X be a set and $S \subseteq X \times X$ be a relation on X. The set of equivalence relations on X is closed under intersection, as one may easily check, and $X \times X$ is an equivalence relation. Thus, the intersection all equivalence relations containing S is the smallest equivalence relation $\sim_S$ containing S. We call $\sim_S$ the equivalence relation generated by S.

Two elements $x, y \in X$ are equivalent under $\sim_S$ if and only if there exist a sequence of elements $z_0, \dots, z_n \in X$ with $x = z_0$ and $y = z_n$ for some $n \geq 1$ such that $z_i = z_{i+1}$, $(z_i, z_{i+1}) \in S$, or $(z_{i+1}, z_i) \in S$ for every $0 \leq i \leq n-1$. To see this, one checks two things: first, that what we have just described defines an equivalence relation on S, and secondly, that any equivalence relation on S must contain every such pair (x, y).

1.3. Binary operations

To give context to the term "binary operation", which we study in this section, here is what one might refer to simply as an "operation".

DEFINITION 1.3.1. A (left) *operation* $\star$ of a set X on a set Y is a function $\star \colon X \times Y \to Y$.

NOTATION 1.3.2. The value $\star(x, y)$ of $(x, y) \in X \times Y$ under an operation $\star \colon X \times Y \to Y$ is denoted $x \star y$. It is often referred to as the product of x and y under $\star$ (when confusion does not arise from this language).

EXAMPLE 1.3.3. The set $\mathbb{R}$ acts on $\mathbb{R}^n$ for any $n \geq 1$ by left diagonal multiplication. That is, we have

$$a \cdot (x_1, \dots, x_n) = (ax_1, \dots, ax_n)$$

for $a \in \mathbb{R}$ and $(x_1, \dots, x_n) \in \mathbb{R}^n$. Geometrically, this is the action of scaling of a vector.

If Z is a subset of Y, we can ask if the values $x \star z$ for $x \in X$ and $z \in Z$ land in Z.

DEFINITION 1.3.4. Let $\star \colon X \times Y \to Y$ be a (left) operation of X on Y. A subset Z of Y is said to be *closed* under $\star$ (or, left multiplication by $\star$) if $x \star z \in Z$ for all $x \in X$ and $z \in Z$.

EXAMPLE 1.3.5. Consider the operation $\cdot\colon \mathbb{Z}\times\mathbb{Z}\to\mathbb{Z}$ of multiplication. The subset E of even numbers in $\mathbb{Z}$ is closed under this operation, which is to say left multiplication by integers. That is, if $a\in\mathbb{Z}$ and $b\in E$, then $ab\in E$. However, the subset O of odd numbers is not closed under this operation. For instance, $2\in\mathbb{Z}$ and $1\in O$, but $2\cdot 1 = 2\notin O$.

DEFINITION 1.3.6. Let $\star\colon X\times Y\to Y$ be an operation and Z be a subset of Y that is closed under $\star$. Then the restriction of $\star$ to an operation of X on Z is the operation $\star_Z\colon X\times Z\to Z$ defined by $x\star_Z z = x\star z$ for all $x\in X$ and $z\in Z$.

In this text, we will most often encounter binary operations.

DEFINITION 1.3.7. A *binary operation* on a set X is an operation of X on itself.

REMARKS 1.3.8. Let X be a set.

a. A binary operation $\star$ on X is simply a function $\star\colon X\times X\to X$.

b. We often refer to a binary operation on X more simply as an “operation” on X, the fact that X is operating on itself being implied.

EXAMPLES 1.3.9. The following are binary operations.

a. Addition (or subtraction) $+$ on $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{R}^n$, and m-by-n matrices $M_{mn}(\mathbb{R})$ with entires in $\mathbb{R}$ for any $m,n\geq 1$.

b. Multiplication $\cdot$ on $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, and square n-by-n matrices $M_n(\mathbb{R})$ for any $n\geq 1$.

c. Composition $\circ$ on the set $\mathrm{Maps}(X,X)$ of maps from a set X to itself, e.g., $X=\mathbb{R}$.

d. Union $\cup$ and intersection $\cap$ on power set $\mathscr{P}_X$ of any set X.

EXAMPLES 1.3.10.

a. Exponentiation is not a binary operation on $\mathbb{C}$, as $(-1)^{1/2}$, for instance, has two possible values. It is therefore not well-defined.

b. Addition is not a binary operation on the set $\mathbb{R}^\times$ of nonzero real numbers, as $-1+1=0$, and $0\notin\mathbb{R}^\times$. We say that $\mathbb{R}^\times$ is not closed under addition.

c. Division in not a binary operation on $\mathbb{R}$, as division by 0 is not defined, but division is a binary operation on $\mathbb{R}^\times$.

We can define a binary operation on a finite set via a multiplication table.

EXAMPLE 1.3.11. Consider the set $X=\{a,b,c\}$. The following table defines a binary operation $\star$ on X:

$\star$	a	b	c
a	b	c	a
b	a	c	c
c	b	a	c

The entry in row b and column a is, by way of example, $b\star a$, and therefore, $b\star a=a$. On the other hand, $a\star b$ is located in the row corresponding to a and column of b, and hence $a\star b=c$.

In the previous example, we could have filled in the nine entries in the bottom right 3-by-3 square arbitrarily with elements of X, as there are no conditions of the values of a binary operation. Often, it is useful to impose conditions that give additional structure.

DEFINITION 1.3.12. Let X be a set.

a. A binary operation $\star$ on X is associative if

$$(x\star y)\star z = x\star(y\star z)$$

for all $x,y,z\in X$.

b. A binary operation $\star$ on X is commutative if

$$x\star y = y\star x$$

for all $x,y\in X$.

EXAMPLES 1.3.13.

a. Addition is associative and commutative on $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{R}^n$, $M_{mn}(\mathbb{R})$, and $\text{Maps}(\mathbb{R},\mathbb{R})$.

b. Subtraction is neither associative nor commutative on the sets $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{R}^n$, $M_{mn}(\mathbb{R})$, and $\text{Maps}(\mathbb{R},\mathbb{R})$.

c. Multiplication is associative and commutative on $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{R}^n$, $\text{Maps}(\mathbb{R},\mathbb{R})$, and is associative but not commutative on $M_n(\mathbb{R})$ for $n\geq 2$.

d. Union and intersection are associative and commutative binary operations on $\mathscr{P}_X$.

e. Composition on $\text{Maps}(X,X)$ is an associative binary operation, but it is not commutative if X has at least 3 elements.

DEFINITION 1.3.14. Let X be a set and $\star$ a binary operation on X. Two elements $x,y\in X$ are said to *commute* under $\star$ if $x\star y = y\star x$.

Commutativity of a binary operation on a finite set can be seen on its mutliplication table, as the table is then symmetric across the diagonal. Associativity is hard to see, but it is a strong condition. Here are some examples.

EXAMPLES 1.3.15. The following are tables of binary operations on the set $\{a,b\}$:

$\star$	a	b
a	b	a
b	b	a

$*$	a	b
a	a	b
b	b	a

$\diamond$	a	b
a	b	a
b	a	a

Of these, only $*$ is associative, while only $*$ and $\diamond$ are commutative, since a and b do not commute under $\star$.

EXAMPLE 1.3.16. We can define operations $+$ and $\cdot$ on $\mathbb{Z}/n\mathbb{Z}$ as follows. Let $a,b\in\mathbb{Z}$, and recall that we denote their equivalence classes under congruence modulo n by $\overline{a},\overline{b}\in\mathbb{Z}/n\mathbb{Z}$. We define $\overline{a}+\overline{b}=\overline{a+b}$ and $\overline{a}\cdot\overline{b}=\overline{a\cdot b}$. These are well-defined, as if c and d are congruent modulo n to a and b, respectively, then $c+d\equiv a+b \bmod n$ and $c\cdot d\equiv a\cdot b \bmod n$.

DEFINITION 1.3.17. A set X together with a binary operation $\star\colon X \times X \to X$ is called a *binary structure*, and we write it as a pair $(X,\star)$.

REMARK 1.3.18. If $(X,\star)$ is a binary structure, then we often refer to X as the *underlying set*.

EXAMPLE 1.3.19. The pair $(\mathbb{Z}/n\mathbb{Z},+)$ is a binary structure, as is $(\mathbb{Z}/n\mathbb{Z},\cdot)$.

DEFINITION 1.3.20. Let $(X,\star)$ be a binary structure. A subset A is said to be *closed* under the binary operation $\star$ if $a\star b \in A$ for all $a,b \in A$.

DEFINITION 1.3.21. Let $(X,\star)$ be a binary structure and A a closed subset of X. Then the *restriction* of $\star$ to A is a binary operation $\star_A\colon A \times A \to A$ defined by $a \star_A b = a\star b$ for all $a,b \in A$. We usually denote $\star_A$ more simply by $\star$.

REMARK 1.3.22. If $(X,\star)$ is a binary operation and A is a closed subset of X, then $(A,\star)$ is a binary structure as well.

EXAMPLES 1.3.23.

a. The subsets $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ of $\mathbb{C}$ are closed under the binary operation $+$.

b. The subset $[-1,1]$ of $\mathbb{R}$ is not closed under $+$, though it is under $\cdot$.

c. The set of all nonempty subsets of a set X is closed under the binary operation $\cup$ on $\mathscr{P}_X$, but not under the operation $\cap$ (unless X has fewer than two elements).

d. The matrices in $M_n(\mathbb{R})$ with determinant 1 are closed under multiplication. The resulting binary structure is denoted $\mathrm{SL}_n(\mathbb{R})$.

REMARK 1.3.24. If $\star\colon X \times X \to X$ is a binary operation, then we can also think of it as an operation. However, the notions of a subset A of X being closed under $\star$ as a binary operation and being closed under $\star$ as an operation do not in general coincide. The first says that $\star$ restricts to a binary operation $\star\colon A \times A \to A$, while the second says that $\star$ restricts to an operation $\star\colon X \times A \to A$. In other words, the first requires only that the product of any two elements of A lands in A (under $\star$), while the second says that the product $x \star a$ lands in A for any $x \in X$ and any $a \in A$, which is a stronger condition.

EXAMPLE 1.3.25. Consider the set $\mathbb{Z}$ and the binary operation $\cdot$ of multiplication on it. The set of odd numbers E is closed under multiplication thought of as a binary operation, since the product of any two odd numbers is odd. However, it is not closed $\cdot$ thought of as an operation of $\mathbb{Z}$ on itself, since the product of an even number and an odd number is not odd (as in Example 1.3.5).

Look for similarities in the following tables of binary structures with underlying sets of order 3.

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\star$	a	b	c
a	a	b	c
b	b	c	a
c	c	a	b

In fact, if we replace $+$ by $\star$, 0 by a, 1 by b, and 2 by c, the first table becomes the second. In a sense, these binary operations are the "same". We give this notation of sameness a technical definition. Note that to be the same in this sense, there must exist a bijection between the sets: i.e., if they are finite, they must have the same number of elements.

DEFINITION 1.3.26. Let $(X,\star)$ and $(Y,*)$ be binary structures. We say that they are *isomorphic* if there exists a bijection $f\colon X \to Y$ such that

$$f(a \star b) = f(a) * f(b)$$

for all $a,b \in X$. We then say that f is an *isomorphism*.

REMARK 1.3.27. If we remove the condition of bijectivity in Definition 1.3.26, then the map $f\colon X \to Y$ with $f(a \star b) = f(a) * f(b)$ is called a *homomorphism of binary structures*.

In the above example $f(0) = a$, $f(1) = b$, and $f(2) = c$, and the condition that $f(x+y) = f(x) \star f(y)$ for all $x,y \in \{0,1,2\}$ is exactly that the multiplication tables match.

EXAMPLES 1.3.28.

a. The map $f\colon \mathbb{Z} \to \mathbb{Z}$ defined by $f(n) = -n$ defines an isomorphism from $(\mathbb{Z},+)$ to itself.

b. The map $f\colon \mathbb{Z} \to \mathbb{Z}$ defined by $f(n) = 2n$ is not an isomorphism from $(\mathbb{Z},+)$ to itself. It satisfies $f(m+n) = f(m) + f(n)$, but it is not onto.

c. Let $\mathbb{R}_{>0} = \{x \in \mathbb{R} \mid x > 0\}$. Define $f\colon \mathbb{R} \to \mathbb{R}_{>0}$ by $f(x) = e^x$. This is an isomorphism from $(\mathbb{R},+)$ to $(\mathbb{R}_{>0},\cdot)$, since

$$e^{x+y} = e^x e^y$$

for all $x,y \in \mathbb{R}$.

d. The map $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ is not an isomorphism from $(\mathbb{R},+)$ to itself, as

$$f(1+1) = 8 \neq 2 = f(1) + f(1).$$

On the other hand, the same map does define an isomorphism from $(\mathbb{R},\cdot)$ to $(\mathbb{R},\cdot)$.

We have the following lemma.

LEMMA 1.3.29. *Suppose that f is an isomorphism from $(X,\star)$ to $(Y,*)$. Then the inverse f^{-1} of f is an isomorphism from $(Y,*)$ to $(X,\star)$.*

PROOF. Let $y_1,y_2 \in Y$. Then there exists $x_1,x_2 \in X$ with $f(x_1) = y_1$ and $f(x_2) = y_2$ We have

$$f^{-1}(y_1) \star f^{-1}(y_2) = x_1 \star x_2,$$

and

$$f(x_1 \star x_2) = f(x_1) * f(x_2) = y_1 * y_2,$$

so

$$x_1 \star x_2 = f^{-1}(y_1 * y_2),$$

as desired. □

EXAMPLE 1.3.30. The inverse of $f\colon \mathbb{R} \to \mathbb{R}_{>0}$ with $f(x) = e^x$ is $f^{-1}(x) = \log(x)$, which satisfies

$$\log(x \cdot y) = \log(x) + \log(y)$$

for $x, y \in \mathbb{R}_{>0}$.

In fact, the properties of being isomorphic puts an equivalence relation on any set of binary structures.

EXAMPLE 1.3.31. The set of representatives for the *isomorphism classes* (i.e., equivalence classes under isomorphism) of binary structures on the set $\{a, b\}$ has 10 elements. That is, one can construct at most 10 tables for binary operations on $\{a, b\}$ that give binary structures, no two of which are isomorphic, as the reader can check.

CHAPTER 2

Group theory

2.1. Groups

In this section, we introduce groups, which can briefly be defined as associative binary structures with identities and inverses. We begin by defining the two latter terms.

DEFINITION 2.1.1. Suppose that $(X,\star)$ is a binary structure.

a. A *left (resp., right) identity element* of X is an element $e \in X$ that satisfies $e \star x = x$ (resp., $x \star e = x$).

b. If $e \in X$ is both a left and a right identity element of X, we say that it is an *identity element* of X.

EXAMPLES 2.1.2.

a. Under addition, 0 is a left and right identity element in $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{R}^n$, $M_n(\mathbb{R})$, and $\mathrm{Maps}(\mathbb{R},\mathbb{R})$, with 0 in the latter three examples being the zero vector, zero matrix, and constant function with value 0. Similarly, under multiplication, 1 is a left and right identity element in all of the latter sets.

b. Under subtraction on the sets from part *a*, the element 0 is a right identity but there is no left identity element.

c. Under composition, $f(x) = x$ is an identity element in $\mathrm{Maps}(\mathbb{R},\mathbb{R})$.

d. Under union, $\varnothing$ is an identity element in $\mathscr{P}_X$.

e. Multiplication is a binary operation on the even integers $2\mathbb{Z}$ but $2\mathbb{Z}$ has no left and no right identity elements.

f. For the binary structure defined on $\{a,b\}$ by the table

$\star$	a	b
a	a	b
b	a	b

a and b are both left identity elements, but there is no right identity element.

One could ask whether or not there can be more than one (left and right) identity element in a binary structure. The following provides the answer.

LEMMA 2.1.3. *Let $(X,\star)$ be a binary structure. Suppose that $e \in X$is a left identity element and that $f \in X$ is a right identity element. Then $e = f$, and in particular e is an identity element in X.*

PROOF. If f is a right identity element, we have $e \star f = e$. On the other hand, since e is a left identity element, we have $e \star f = f$. Therefore, we have $e = f$. □

The following is an immediate corollary.

COROLLARY 2.1.4. *Let $(X, \star)$ be a binary structure that contains an identity element e. Then every (left or right) identity element in X is equal to e.*

DEFINITION 2.1.5. Suppose that $(X, \star)$ is a binary structure with an identity element $e \in X$.

a. A *left (resp., right) inverse* of $x \in X$ is an element $y \in X$ such that $y \star x = e$ (resp., $x \star y = e$).

b. An element that is both a left and a right inverse to $x \in X$ is called an *inverse* of $x \in X$.

EXAMPLES 2.1.6.

a. In $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $M_n(\mathbb{R})$, and $\text{Maps}(\mathbb{R}, \mathbb{R})$, the negative $-x$ of an element called x is the inverse under addition. Under multiplication, $x^{-1} = 1/x$ is the inverse of any $x \neq 0$ in $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$. The elements that have multiplicative inverses in $\mathbb{Z}$ are ± 1, in $M_n(\mathbb{R})$ they are the matrices with nonzero determinant, and in $\text{Maps}(\mathbb{R}, \mathbb{R})$ they are the nowhere vanishing functions.

b. Under subtraction on the sets of part a, an element x is its own left and right inverse.

c. Under composition, an element $f \in \text{Maps}(\mathbb{R}, \mathbb{R})$ has an inverse f^{-1} if and only if it is a bijection.

d. Under union on $\mathscr{P}_X$, only $\varnothing$ has an inverse, which is itself.

e. For the binary structure defined on $\{a, b, c\}$ by the table

$\star$	a	b	c
a	a	b	c
b	b	a	a
c	c	b	c

a is an identity element and is its own inverse, b is an inverse of itself, c is a right inverse of b and therefore b is a left inverse of c, but c has no right inverse.

LEMMA 2.1.7. *Let $(X, \star)$ be a binary structure with an identity element e. Suppose that $x \in X$ has a left inverse y and a right inverse z. Then $y = z$.*

PROOF. We need only write down the chain of equalities

$$y = y \star e = y \star (x \star z) = (y \star x) \star z = e \star z = z.$$

□

With the concepts of identity elements and inverses in hand, we now give the full definition of a group.

DEFINITION 2.1.8. A *group* is a set G together with a binary operation $\star \colon G \times G \to G$ such that

i. $\star$ is associative,

ii. there exists an element $e \in G$ such that $e \star x = x = x \star e$, and

iii. for every $x \in G$, there exists an element $y \in G$ such that $x \star y = e = y \star x$.

In other words, a group is a set with an associative binary operation, an identity element, and inverses with respect to that identity element.

Here are some examples of groups.

EXAMPLES 2.1.9.

a. Under addition, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $M_n(\mathbb{R})$, and $\mathrm{Maps}(\mathbb{R},\mathbb{R})$ are all groups.

b. For $X = \mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$, we set $X^\times = X - \{0\}$. Under multiplication, $\mathbb{Q}^\times$, $\mathbb{R}^\times$, and $\mathbb{C}^\times$ are groups.

c. Under multiplication, the set $\mathrm{GL}_n(\mathbb{R})$ of invertible n by n-matrices (i.e., those with nonzero determinant) forms a group, known as the *general linear group*.

d. Under multiplication, the set of nowhere vanishing functions in $\mathrm{Maps}(\mathbb{R},\mathbb{R})$ forms a group.

e. The set $\{e\}$ consisting of a single element is a group under the binary operation $\star$ defined by $e \star e = e$. This group is known as the *trivial group*.

On the other hand, here are some of many binary structures that are not groups.

EXAMPLES 2.1.10.

a. The integers are not a group under multiplication, nor are $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$ before removing 0.

b. The set $\mathrm{Maps}(\mathbb{R},\mathbb{R})$ is not a group under composition, as not every function has an inverse.

c. The set $\mathscr{P}_X$ of subsets of a set is not a group under union.

The following theorem is used in showing the uniqueness of inverses.

PROPOSITION 2.1.11 (Cancellation theorem). *Let G be a group, and let $x, y, z \in G$ be such that*

$$x \star y = x \star z \qquad (\textit{resp., } y \star x = z \star x).$$

Then $y = z$.

PROOF. We prove the first statement. Let x' be any (left) inverse to x. Under the given assumption, we have

$$y = e \star y = (x' \star x) \star z = x' \star (x \star y) = x' \star (x \star z) = (x' \star x) \star z = e \star z = z.$$

□

The following is now quickly derived.

LEMMA 2.1.12. *Let G be a group. If $y, z \in G$ are both inverses to $x \in G$ on either the left or the right (or both), then $y = z$.*

PROOF. Suppose first that y and z are right inverses to x. Then we have

$$x \star y = e = x \star z,$$

and the result now follows from the cancellation theorem. A similar argument holds if both y and z are right inverses. In fact, even if y is a left inverse and z is a right inverse, there is by definition of the group a third element x' in the group that is both a left an a right inverse, and so equals both y and z by what we have just proven. So y and z must be equal. □

NOTATION 2.1.13. Let G be a group and $x \in G$ an element. Suppose the operation on G is not denoted $+$. Then we (almost invariably) use the following notation.

a. The unique inverse to x is written x^{-1}.

b. Let $n \in \mathbb{Z}$. We set $x^0 = e$. If $n \geq 1$, we usually write x^n for $x \star x \star \cdots \star x$, the product being of n copies of x, which is unambiguously defined by the associativity of $\star$.

If the binary operation on the group is denoted $+$, then we write the inverse of x as $-x$ and nx instead of x^n.

REMARK 2.1.14. Let G be a group and e an element for which the operation is not denoted as $+$. The reader should be able to check that for $x \in G$ and $m,n \in \mathbb{Z}$, one has

$$x^{m+n} = x^m x^n, \qquad x^{mn} = (x^m)^n, \qquad x^n = (x^{-1})^{-n} = (x^{-n})^{-1}, \qquad \text{and } e^n = e.$$

DEFINITION 2.1.15. Let G be a group.

a. We say that G is *abelian* if its binary operation is commutative.

b. We say that G is *nonabelian* if its binary operation is not commutative.

EXAMPLES 2.1.16.

a. All of $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $M_{mn}(\mathbb{R})$, and $\mathrm{Maps}(\mathbb{R},\mathbb{R})$ are abelian groups under addition.

b. The groups $\mathbb{Q}^\times$, $\mathbb{R}^\times$, and $\mathbb{C}^\times$ are abelian (under multiplication).

c. The group $\mathrm{GL}_n(\mathbb{R})$ is nonabelian if $n \geq 2$.

REMARK 2.1.17. From now on, we will drop the use of $\star$ for an arbitrary binary operation, and simply use the more conventional symbol $\cdot$. However, the reader should keep in mind that this does not mean that the operation in question is multiplication. Moreover, we shall often write $x \cdot y$ more simply as xy.

LEMMA 2.1.18. *Let G be a group. For $x,y \in G$, we have $(xy)^{-1} = y^{-1}x^{-1}$.*

PROOF. We have

$$(y^{-1}x^{-1})(xy) = y^{-1}(x^{-1}(xy)) = y^{-1}((x^{-1}x)y) = y^{-1}(ey) = y^{-1}y = e.$$

Therefore $y^{-1}x^{-1}$ is left inverse to xy, and so by Lemma 2.1.12 it equals $(xy)^{-1}$. □

We end this section with a few more examples of groups.

EXAMPLE 2.1.19. The set $\mathbb{Z}/n\mathbb{Z}$ of congruence classes modulo n forms a group under the addition law

$$\overline{a} + \overline{b} = \overline{a+b}.$$

The identity is $\overline{0}$, and the inverse of $\overline{a}$ is $\overline{-a}$.

Clearly, $\mathbb{Z}/n\mathbb{Z}$ is an abelian group.

REMARK 2.1.20. Usually, we simply write a for $\overline{a}$. We have kept up the distinction to this point to make clear the difference between a and its equivalence class. From now on, however, if we understand that we are working in $\mathbb{Z}/7\mathbb{Z}$, e.g., from context, we will write equations such as $5+2=0$, with the fact that we are working with equivalence classes as above being understood.

DEFINITION 2.1.21. The *symmetric group S_X* on a set X is the set

$$S_X = \{f \colon X \to X \mid f \text{ is bijective}\}$$

with the binary operation $\circ$ of composition.

DEFINITION 2.1.22. Let X be a set. An element of S_X is referred to as a *permutation* of X. We say that $\sigma \in S_X$ *permutes* the elements of X.

REMARK 2.1.23. The group S_X is alternately referred to as the *group of permutations* of a set X.

REMARK 2.1.24. The group S_X is nonabelian if X has at least three elements.

EXAMPLE 2.1.25. If $X = \mathbb{R}$, then $f(x) = x+1$ and $g(x) = x^3$ both lie in $S_{\mathbb{R}}$, but do not commute.

DEFINITION 2.1.26. When $X = \{1,2,\ldots,n\}$, then we set $S_n = S_X$, and we refer to S_n as the *symmetric group on n letters*.

REMARK 2.1.27. The notion of isomorphism of binary structures carries over to groups. An *isomorphism* of groups is just an isomorphism of the underlying binary structures, i.e., a bijection $f \colon G \to G'$ between groups G and G' such that

$$f(x \cdot y) = f(x) \cdot f(y)$$

for each $x, y \in G$. If G and G' are isomorphic, we write $G \cong G'$ (noting that the property of being isomorphic forms an equivalence class on any set of groups).

EXAMPLES 2.1.28.

a. The group $\mathrm{GL}_1(\mathbb{R})$ is isomorphic to $\mathbb{R}^\times$ via the map $f \colon \mathbb{R}^\times \to \mathrm{GL}_1(\mathbb{R})$ defined by $f(a) = (a)$.

b. Let X be a set with exactly n elements, say $X = \{x_1, x_2, \ldots, x_n\}$. Then we define an isomorphism

$$f \colon S_n \xrightarrow{\sim} S_X, \qquad f(\sigma)(x_i) = x_{\sigma(i)},$$

which is to say that f takes a permutation $\sigma \in S_n$ that takes i to some other number j to the permutation in S_X that maps x_i to x_j. In other words, it doesn't matter whether we're permuting n cars or n apples: the groups are isomorphic.

To every group, we have an associated opposite group.

DEFINITION 2.1.29. The *opposite group G^{op}* of a group G is the set G together with the operation $x \star y = yx$ for $x, y \in G$.

EXAMPLE 2.1.30. The opposite group of an abelian group is the original group.

2.2. Subgroups

DEFINITION 2.2.1. A subset H of a group G is a *subgroup* if it is closed under the binary operation on G and is a group with respect to the restriction of that operation to a binary operation on H. If H is a subgroup of G, we write $H \leqslant G$.

More succinctly, a subset of a group is a subgroup if it is a group with respect to the operation on the group.

REMARK 2.2.2. The relation $\leqslant$ is a partial ordering on the set of subgroups of a group.

DEFINITION 2.2.3.

a. The set $\{e\}$ containing only the identity element of G is a subgroup of G known as the *trivial subgroup* (as it is a trivial group that is also a subgroup).

b. A subgroup H of G that is not the trivial subgroup is called *nontrivial*.

DEFINITION 2.2.4. If H is a subgroup of G with $H \neq G$, then we say that H is a *proper subgroup* of G, and we write $H < G$.

EXAMPLES 2.2.5. The groups $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ under addition are all subgroups of $\mathbb{C}$.

To check that a group is a subgroup, one usually employs the following criteria.

THEOREM 2.2.6. *A subset H of a group G is a subgroup under the restriction on the binary operation $\cdot$ on G if and only if*

(0) $e \in H$,

(1) H is closed under $\cdot$,

(2) if $h \in H$, then $h^{-1} \in H$.

PROOF. If H is a subgroup of G with respect to $\cdot$, then it is by definition closed under $\cdot$. Since H is a group under $\cdot$, there exists an element $f \in H$ with $f \cdot h = h$ for all $h \in H$. By the cancellation theorem, we then have $f = e$, so $e \in H$. Also, for each $h \in H$, we have an element $h' \in H$ with $h \cdot h' = e$. As $e = h \cdot h^{-1}$, the cancellation theorem again tells us that $h' = h^{-1}$, so $h^{-1} \in H$. Therefore, the conditions (0)-(2) hold.

Conversely, if conditions (0)-(2) hold, then H is a binary structure under $\cdot$ by (1) and (0) and (2) leave us only to verify associativity in the definition of a group. However, this follows automatically on H from the associativity of $\cdot$ on the larger set G. □

EXAMPLES 2.2.7. The subset $2\mathbb{Z}$ of $\mathbb{Z}$ is a subgroup under $+$. To see this, note that 0 is even, the sum of two even integers is even, and the negative of an even integer is also even.

EXAMPLE 2.2.8. The subset

$$\mathrm{SL}_n(\mathbb{R}) = \{A \in \mathrm{GL}_n(\mathbb{R}) \mid \det(A) = 1\}$$

of $\mathrm{GL}_n(\mathbb{R})$ is a subgroup under $\cdot$, known as the *special linear group*. We use Theorem 2.2.6 to check this:

(0) We have $\det I_n = 1$, so $I_n \in \mathrm{SL}_n(\mathbb{R})$.

(1) If $A, B \in \mathrm{SL}_n(\mathbb{R})$, then

$$\det(A \cdot B) = \det(A) \cdot \det(B) = 1$$

so $A \cdot B \in \mathrm{SL}_n(\mathbb{R})$.

(2) If $A \in \mathrm{SL}_n(R)$, then

$$\det(A^{-1}) = \det(A)^{-1} = 1$$

so $A^{-1} \in \mathrm{SL}_n(\mathbb{R})$.

EXAMPLE 2.2.9. Let

$$S^1 = \{z \in \mathbb{C} \mid |z| = 1\} = \{e^{2\pi i\theta} \mid \theta \in \mathbb{R}\}.$$

Here $e^{2\pi i\theta}$ corresponds to the point $(\cos\theta, \sin\theta)$ on the unit circle in the usual model of the complex plane. In fact, recall that

$$e^{2\pi i\theta} = \cos\theta + i\sin\theta \in \mathbb{C}$$

and

$$|\cos\theta + i\sin\theta| = \sqrt{\cos^2(\theta) + \sin^2(\theta)} = 1.$$

Then S^1 is a subgroup of $\mathbb{C}^\times$ under $\cdot$. To see this, we check:

(0) We have $|1| = 1$.

(1) If $z, w \in S^1$, then $z = e^{2\pi i\theta}$ and $w = e^{2\pi i\psi}$ for some $\theta, \psi \in \mathbb{R}$. We have

$$zw = e^{2\pi i(\theta+\psi)} \in S^1.$$

(2) If $z = e^{2\pi i\theta}$, then

$$z^{-1} = e^{2\pi i(-\theta)} \in S^1.$$

Theorem 2.2.6 has the following shorter formulation.

COROLLARY 2.2.10. *A nonempty subset H of a group is a subgroup under the restriction of the binary operation $\cdot$ on G if and only if $h \cdot k^{-1} \in H$ for all $h, k \in H$.*

PROOF. If H is a subgroup of G and $h, k \in H$, then $k^{-1} \in H$ and, consequently, $hk^{-1} \in H$ by Theorem 2.2.6. Conversely, suppose $hk^{-1} \in H$ for all $h, k \in H$. As H is nonempty, let $h \in H$. Using this critersion, we have successively that $e = hh^{-1} \in H$, $k^{-1} = ek^{-1} \in H$, and $hk = h(k^{-1})^{-1} \in H$, so Theorem 2.2.6 implies that H is a subgroup. □

DEFINITION 2.2.11.

a. A group G is *finite* if its underlying set is finite. Otherwise, we say that G is *infinite*.

b. The *order* $|G|$ of a finite group G is the order (number of elements in) of the underlying set. If G is infinite, we say that its order is *infinite*.

The following provides an interesting example.

LEMMA 2.2.12. *Let $n \geq 1$. The group S_n is finite of order $n!$.*

PROOF. For an arbitrary element $\sigma \in S_n$, we have n choices for the value $\sigma(1)$. Then $\sigma(2)$ can be any of the remaining $n-1$ values, and $\sigma(3)$ is one of the then remaining $n-2$ values, and so forth, until one value is left for $\sigma(n)$. Therefore, the order of S_n is $n \cdot (n-1) \cdots 1 = n!$. □

EXAMPLE 2.2.13. As a further subgroup of S^1 (so also a subgroup of $\mathbb{C}^\times$), we have

$$\mu_n = \{z \in \mathbb{C}^\times \mid z^n = 1\} = \{e^{2\pi ik/n} \mid k \in \mathbb{Z}\}.$$

To see the equality of the latter two sets, note that $(e^{2\pi ik/n})^n = 1$. On the other hand $z^n = 1$ implies that $|z|^n = 1$, so $|z| = 1$, which means that $z = e^{2\pi i\theta}$ for some $\theta \in \mathbb{R}$. But the only way that $(e^{2\pi i\theta})^n = 1$ can hold is for $n\theta$ to be an integer, which means exactly that $\theta = k/n$ for some $n \in \mathbb{Z}$. Note that $|\mu_n| = n$, since $e^{2\pi ik/n} = e^{2\pi ij/n}$ if and only if $j \equiv k \bmod n$. That this order equals $|\mathbb{Z}/n\mathbb{Z}| = n$ is no coincidence. In fact, these two groups are isomorphic, as well shall see in the following section.

2.3. Cyclic groups

DEFINITION 2.3.1. Let G be a group, and let $g \in G$. The *cyclic subgroup* of G *generated by* g is the group

$$\langle g \rangle = \{g^n \mid n \in \mathbb{Z}\}.$$

LEMMA 2.3.2. *Let $g \in G$. Then $\langle g \rangle$ is the smallest subgroup of G containing g.*

PROOF. Since the smallest subgroup of G containing G is itself a group, it must contain g^n for all $n \in \mathbb{Z}$, so it contains $\langle g \rangle$. On the other hand, we see that $\langle g \rangle$ is a subgroup of G since it contains $e = g^0$, is closed under multiplication (as $g^m \cdot g^n = g^{m+n}$), and contains inverses (as $(g^n)^{-1} = g^{-n}$). Being that $\langle g \rangle$ is a subgroup of G contained in the smallest subgroup containing g, it is itself the smallest subgroup. □

EXAMPLES 2.3.3.

a. The cyclic subgroup $\langle 2 \rangle$ of $\mathbb{Z}$ generated by 2 is $2\mathbb{Z}$.

b. The cyclic subgroup of $\mathrm{GL}_2(\mathbb{R})$ generated by

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

is

$$\langle A \rangle = \{I_2, A, -I_2, -A\}.$$

DEFINITION 2.3.4.

a. A group G is called *cyclic* if there exists $g \in G$ with $G = \langle g \rangle$.

b. An element of g of a group G is called a *generator* if $G = \langle g \rangle$. We then say that g *generates* G and that G is *generated* by g.

REMARK 2.3.5. Of course, any cyclic subgroup of a group G is itself a cyclic group.

EXAMPLES 2.3.6.

a. The group $\mathbb{Z}$ is cyclic, generated by 1.

b. The group $\mathbb{Z}/n\mathbb{Z}$ is cyclic for any $n \geq 0$, again generated by 1.

c. The group μ_n is cyclic with generator $e^{2\pi i/n}$.

d. The trivial group is a cyclic group of order 1.

REMARK 2.3.7. Every cyclic group is abelian, since powers of a generator commute.

DEFINITION 2.3.8. Let G be a group. The *order* of an element $g \in G$ is the smallest positive integer n such that $g^n = e$, if it exists. If such an n exists, then g is said to have *finite order*, and otherwise g is said to have *infinite order*.

PROPOSITION 2.3.9. *Let g be an element in a group. Then the order of $\langle g \rangle$ and the order of g are equal if either is finite (and both infinite otherwise). Moreover, for any $i, j \in \mathbb{Z}$, we have $g^i = g^j$ if and only if*

- $i \equiv j \bmod n$*, if g is finite of order n, and*
- $i = j$*, if g has infinite order.*

PROOF. First, suppose that g has finite order n. If $g^i = g^j$, then $g^{i-j} = e$. Note that $g^n = e$ as well. Dividing $i - j$ by n, we have

$$i - j = qn + r$$

for some quotient $q \in \mathbb{Z}$ and remainder $0 \leq r \leq n-1$. We then have

$$e = g^{i-j} = g^{qn+r} = (g^n)^q g^r = g^r,$$

but $r < n$ and n is minimal, so $r = 0$. That is, $i - j$ is a multiple of n, so $i \equiv j \bmod n$. In particular, the distinct elements of $\langle g \rangle$ are exactly $e, g, \ldots, g^{n-1}$, so $\langle g \rangle$ has order n.

If g has infinite order, then for $g^i = g^j$ to hold, one must have $g^{i-j} = e$, which forces $i = j$. Therefore, all of the powers of g are distinct, and $\langle g \rangle$ is infinite. □

LEMMA 2.3.10. *Suppose that G is a cyclic group. If G is infinite, then G is isomorphic to $\mathbb{Z}$. Otherwise, G is isomorphic to $\mathbb{Z}/n\mathbb{Z}$, where $n = |G|$.*

PROOF. Let g be a generator of G. Suppose first that G is infinite. We define a map

$$f \colon \mathbb{Z} \to G, \qquad f(i) = g^i \text{ for all } i \in \mathbb{Z}.$$

This is one-to-one since $f(i) = f(j)$ implies $g^i = g^j$, which can only happen if $i = j$ by Proposition 2.3.9. It is onto as every element of $\langle g \rangle$ has the form $g^i = f(i)$ for some i. It is then an isomorphism of groups as

$$f(i + j) = g^{i+j} = g^i g^j = f(i)f(j).$$

If $|G| = n$, then we define

$$f \colon \mathbb{Z}/n\mathbb{Z} \to G, \qquad f(i) = g^i \text{ for all } i \in \mathbb{Z}.$$

This is well-defined as $f(i + qn) = g^{i+qn} = g^i$, so it is independent of the choice of representative of i modulo n. It is one-to-one as $g^i = g^j$ implies $i = j$ in $\mathbb{Z}/n\mathbb{Z}$ by Proposition 2.3.9. It is then onto and an isomorphism for the same reasons as in the infinite case. □

As a result, the group μ_n is isomorphic to $\mathbb{Z}/n\mathbb{Z}$ (under the map taking $e^{2\pi ik/n}$ to k). The groups $n\mathbb{Z}$ for $n \geq 1$ are all isomorphic to $\mathbb{Z}$ itself, but for this one must take the map $\mathbb{Z} \to n\mathbb{Z}$ that is multiplcation by n.

THEOREM 2.3.11. *Every subgroup of a cyclic group is cyclic.*

PROOF. Let H be a subgroup of a cyclic group G with generator g. Let $k \geq 1$ be minimal such that $g^k \in H$. We claim that $H = \langle g^k \rangle$. Since H is closed under multiplication and inverses, it must contain every power of g^k, so it contains the subgroup $\langle g^l \rangle$. Now suppose that $g^i \in H$ for some $i \in \mathbb{Z}$. Again, divide i by k and get $q \in \mathbb{Z}$ and $0 \leq r \leq k-1$ with $i = qk + r$. Then $g^i = (g^k)^q g^r$, so

$$g^r = g^i (g^k)^{-q} \in H,$$

in that H is a subgroup. But minimality forces $r = 0$, so i is a multiple of k, proving the claim. □

COROLLARY 2.3.12. *The subgroups of $\mathbb{Z}$ are exactly the $n\mathbb{Z} = \langle n \rangle$ with n a nonnegative integer.*

Let's consider the subgroups of $\mathbb{Z}/n\mathbb{Z}$ for some $n \geq 1$, which we now know to be cyclic. Recall that the greatest common divisor $\gcd(i,j)$ of two integers i and j that are not both zero is defined to be the smallest positive integer dividing both i and j. We also set $\gcd(0,0) = 0$.

LEMMA 2.3.13. *Given $i, j \in \mathbb{Z}$, we have*

$$\langle \gcd(i,j) \rangle = \{ai + bj \mid a, b \in \mathbb{Z}\}.$$

PROOF. In the case that $i = j = 0$, we have that both sides equal $\langle 0 \rangle$, so the lemma holds, and therefore we may assume that at least one is nonzero. Since $\gcd(i,j)$ divides both i and j, we have $i, j \in \langle \gcd(i,j) \rangle$. As a subgroup, the latter group is closed under addition and taking of negatives, so $ai + bj$ is in it as well. In other words, $H = \{ai + bj \mid a, b \in \mathbb{Z}\}$ is contained in $\langle \gcd(i,j) \rangle$.

Conversely, note that the set H is a (nontrivial) subgroup of $\mathbb{Z}$ in that it satisfies all of the properties of one, so it equals $\langle d \rangle$ for some $d \geq 1$. Since $i, j \in \langle d \rangle$ by definition, we have that d divides both i and j, and therefore is less than or equal to $\gcd(i,j)$. On the other hand, we know that $d \in \langle \gcd(i,j) \rangle$, so $\gcd(i,j) \leq d$, and therefore $d = \gcd(i,j)$. In other words, we have $H = \langle \gcd(i,j) \rangle$. □

PROPOSITION 2.3.14. *Every subgroup of $\mathbb{Z}/n\mathbb{Z}$ has the form $\langle d \rangle$ for some $d \geq 1$ dividing n. In fact, for any $j \in \mathbb{Z}$, we have $\langle j \rangle = \langle \gcd(j,n) \rangle$.*

PROOF. The second statement implies the first, so we focus on it. Since $\gcd(j,n)$ divides j, we have that $\langle j \rangle \leqslant \langle \gcd(j,n) \rangle$. On the other hand, we have by Lemma 2.3.13 that

$$\gcd(j,n) \in \{aj + bn \mid a, b \in \mathbb{Z}\}$$

inside $\mathbb{Z}$, which means that $\gcd(j,n) \equiv aj \bmod n$ for some $a \in \mathbb{Z}$. In other words, in $\mathbb{Z}/n\mathbb{Z}$, we have $\gcd(j,n) \in \langle j \rangle$, so $\langle \gcd(j,n) \rangle \leqslant \langle j \rangle$, as desired. □

REMARK 2.3.15. The subgroup $\langle n \rangle$ of $\mathbb{Z}/n\mathbb{Z}$ is just the trivial subgroup $\langle 0 \rangle = \{0\}$.

Recall that two integers are said to be relatively prime if their greatest common divisor is 1.

COROLLARY 2.3.16. *Let G be a group and $g \in G$ an element of order n.*

a. For $i \in \mathbb{Z}$, the order of g^i is n/d, where $d = \gcd(i,n)$, and $\langle g^i \rangle = \langle g^d \rangle$.

b. The generators of $\langle g \rangle$ are the g^i with i relatively prime to n.

PROOF. Consider the isomorphism $\phi \colon G \to \mathbb{Z}/n\mathbb{Z}$ under which g^i is taken to i. This carries the subgroup $\langle g^i \rangle$ bijectively to the subgroup $\langle i \rangle$, which by Proposition 2.3.14 equals $\langle d \rangle$. But the latter group has elements $0, d, 2d, \ldots, (n/d-1)d$, so has order n/d. As ϕ is a bijection, part a is then seen to hold. Part b then follows immediately from part a, as the i for which $\langle g^i \rangle = \langle g \rangle$ are the i with $\gcd(i,n) = 1$. □

DEFINITION 2.3.17. The *Euler phi-function* is the map $\varphi \colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ such that $\varphi(n)$ is the number of relatively prime integers to n between 1 and n.

REMARK 2.3.18. The Euler phi-function φ has the properties that $\varphi(mn) = \varphi(m)\varphi(n)$ whenever $\gcd(m,n) = 1$ and that $\varphi(p^r) = p^{r-1}(p-1)$ for a prime number p and $r \geq 1$. Its values on $1,2,3,4,5,\ldots$ are $1,1,2,2,4,2,6,4,6,4,10,4,12,6,8,8,16,\ldots$.

REMARK 2.3.19. It follows from Corollary 2b that the number of generators of a cyclic group G of order n is exactly $\varphi(n)$, where φ is the Euler phi-function.

2.4. Generators

The relation $\leqslant$ is a partial ordering on any set of subgroups of a group. The following proposition asserts the existence of minimal elements of certain such subsets. It is a consequence of Lemma 1.2.24, but prove it here for convenience.

PROPOSITION 2.4.1. *Let G be a group, and let S be a nonempty subset of G. Then there exists a smallest subgroup $\langle S \rangle$ of G containing S.*

PROOF. The set P_S of subgroups of G containing S is nonempty, for it contains G itself. Set

$$\langle S \rangle = \bigcap_{H \in X} H.$$

As each $H \in X$ contains G, so does $\langle S \rangle$. Moreover, an arbitrary intersection of subgroups of G is easily verified to itself be a subgroup of G, so $\langle S \rangle$ is a subgroup. Finally, if H is any subgroup of G containing S, then $H \in X$, so $\langle S \rangle \leqslant H$ by definition of the intersection, so $\langle S \rangle$ is the smallest such subgroup (i.e., the unique minimal element of X). □

DEFINITION 2.4.2. The smallest subgroup $\langle S \rangle$ containing a set S is the *subgroup of G generated by S*.

While this definition is rather abstract, we do have the following more concrete description of the elements of $\langle S \rangle$.

PROPOSITION 2.4.3. *Let S be a nonempty subset of G. An element $g \in G$ is contained in $\langle S \rangle$ if and only if g may be written as a product of powers of elements of S: i.e.,*

$$g = s_1^{m_1} s_2^{m_2} \cdots s_k^{m_k}$$

for some $k \geq 0$, $s_i \in S$ and $m_i \in \mathbb{Z}$ for $1 \leq i \leq k$.

PROOF. First, let g be an element that is a product of powers of elements of S. Since $\langle S\rangle$ is a subgroup, it is closed under integer powers and products, so $g \in \langle S\rangle$.

Conversely, note that the set H of elements that are products of powers of elements of S is a subgroup of G, as it contains $e = s^0$ for $s \in S$, is closed under products by definition, and is closed under inverses as

$$(s_1^{m_1} s_2^{m_2} \cdots s_k^{m_k})^{-1} = s_k^{-m_k} \cdots s_2^{-m_2} s_1^{-m_1}.$$

As H is a subgroup of G containing S but contained in $\langle S\rangle$ and $\langle S\rangle$ is the minimal such subgroup, we have $H = \langle S\rangle$. Thus, any element of $\langle S\rangle$ may be written as a product of powers of elements of S, as desired. □

DEFINITION 2.4.4. We say that a subset S of G *generates* G if $G = \langle S\rangle$, and then S is said to be a set of generators of G.

DEFINITION 2.4.5. We say that a group G is *finitely generated* if there exists a finite set of generators of G.

REMARK 2.4.6. If G can be generated by a finite set $\{g_1, g_2, \dots, g_n\}$, we usually write

$$\langle g_1, g_2, \dots, g_n\rangle$$

instead of

$$\langle \{g_1, g_2, \dots, g_n\}\rangle,$$

and we say that G is generated by $g_1, g_2, \dots, g_n$.

EXAMPLE 2.4.7. A cyclic group is finitely generated: in fact, it is generated by a single element.

EXAMPLE 2.4.8. Any finite group is finitely generated, as it is generated by itself.

EXAMPLE 2.4.9. Consider the subgroup G of $\mathrm{GL}_2(\mathbb{R})$ that is

$$G = \left\{ \begin{pmatrix} (-1)^i & b \\ 0 & (-1)^j \end{pmatrix} \middle| i, j, b \in \mathbb{Z} \right\}.$$

It can be generated by the set

$$\left\{ \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right\}.$$

To see this, note that

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^b = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix},$$

and

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}^j \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}^i = \begin{pmatrix} (-1)^i & b \\ 0 & (-1)^j \end{pmatrix}.$$

The group G is not cyclic as it is infinite but contains elements of order 2, but all infinite cyclic groups are isomorphic to $\mathbb{Z}$. In fact, G cannot be generated by any two of its elements: the proof of this more tricky fact is left to the reader.

EXAMPLE 2.4.10. The group $\mathbb{Q}$ can be generated by the set $\{\frac{1}{n} \mid n \geq 1\}$. However, $\mathbb{Q}$ is not finitely generated. For, any integer $N \geq 1$ and nonzero integers a_i, b_i with $b_i > 0$ for $1 \leq i \leq N$. Any element of

$$\left\langle \frac{a_1}{b_1}, \frac{a_2}{b_2}, \ldots, \frac{a_N}{b_N} \right\rangle$$

must have denominator, when put in reduced form, that is a divisor of $b_1 b_2 \cdots b_N$. But clearly not every fraction has such a denominator, so $\mathbb{Q}$ cannot be finitely generated.

2.5. Direct products

Given any two groups, we can form a new group out of them, known as the direct product, whose underlying set is in fact exactly the direct product of the underlying sets of the groups in question.

DEFINITION 2.5.1. Let G and G' be groups. The *direct product* of G and G' is the binary structure $G \times G'$ that is the direct product of the sets G and G' together with the binary operation defined by

$$(a,a') \cdot (b,b') = (a \cdot b, a' \cdot b')$$

for $a,b \in G$ and $a',b' \in G'$.

One might expect the direct product of G and G' to be a group, and in fact it is. The straightforward check is left to the reader.

LEMMA 2.5.2. *The direct product $G \times G'$ of two groups is a group.*

Of course, using this construction, we can think up more examples of new groups than we can mention, e.g., $S_m \times \mathrm{GL}_n(\mathbb{R})$ for any $m,n \geq 1$. The following remarks are easily verified from the definition of the direct product.

REMARK 2.5.3. The group $G \times G'$ is abelian if and only if both G and G' are abelian.

REMARK 2.5.4. If $f \colon G \to H$ is a group isomorphism and G' is another group, then the map

$$f' \colon G \times G' \to H \times G'$$

given by $f'(g,g') = (f(g),g')$ for $g \in G$ and $g' \in G'$ is an isomorphism as well.

REMARK 2.5.5. Direct product forms an associative and commutative binary operation on any set of isomorphism classes of groups. That is, for any groups G_1, G_2, and G_3, we have

$$(G_1 \times G_2) \times G_3 \cong G_1 \times (G_2 \times G_3) \quad \text{and} \quad G_1 \times G_2 \cong G_2 \times G_1.$$

In particular, the associativity means it makes sense to speak of the group

$$G_1 \times G_2 \times \cdots \times G_n$$

for any groups G_1, G_2, $\ldots G_n$.

REMARK 2.5.6. If each of the groups $G_1, G_2, \ldots, G_n$ is finite, then

$$|G_1 \times G_2 \times \cdots \times G_n| = \prod_{i=1}^{n} |G_i|.$$

NOTATION 2.5.7. We write G^n for the direct product $G \times G \times \cdots \times G$ of n copies of G.

REMARK 2.5.8. More generally, for any collection

$$\{G_i \mid i \in I\}$$

of groups G_i for i in some indexing set I, we can put a binary operation on the direct product set

$$\prod_{i \in I} G_i$$

given by coordinate-wise multiplication

$$(a_i)_{i\in I} \cdot (b_i)_{i\in I} = (a_i \cdot b_i)_{i\in I},$$

and the resulting group is known as the *direct product* of the G_i.

Let $n \geq 1$, and let G_i be a group for each $1 \leq i \leq n$. Let

$$G = G_1 \times G_2 \times \cdots \times G_n.$$

For $g \in G_i$, let $g^{(i)} \in G$ denote the element

$$g^{(i)} = (e, \ldots, e, g, e, \ldots, e) \in G$$

that is nontrivial in only the ith coordinate of G and g in the ith coordinate.

PROPOSITION 2.5.9. *Suppose that S_i is a generating set of G_i for each $1 \leq i \leq n$. Then*

$$S = \bigcup_{i=1}^{n} \{g^{(i)} \mid g \in S_i\}$$

is a generating set of G.

PROOF. Suppose $g_i \in G_i$ for each $1 \leq i \leq n$. Then

$$(g_1, g_2, \ldots, g_n) = g_1^{(1)} g_2^{(2)} \cdots g_n^{(n)} \in \langle S \rangle.$$

□

For example, if each G_i is cyclic with generator g_i, then the set $\{g_i^{(i)} \mid 1 \leq i \leq n\}$ generates G. While it is immediate from Proposition 2.5.9 that finite direct products of finitely generated groups are finitely generated, infinite direct products of nontrivial groups are never finitely generated.

EXAMPLE 2.5.10. The group

$$G = \prod_{i=1}^{\infty} (\mathbb{Z}/2\mathbb{Z})$$

is not finitely generated. We give a very brief sketch of the proof: one checks that any finite set of elements X in G must have the property that there exist positive integers j and k such that for each $x = (x_i) \in X$, we have $x_j = x_k$. Then every element in $\langle X \rangle$ has this property, and since not every element of G has this property, we have $\langle X \rangle \neq G$.

The following result gives a general recipe for determining the order of $(g_1, g_2, \ldots, g_n)$.

THEOREM 2.5.11. *Suppose that $g_i \in G_i$ for each $1 \le i \le n$. The order of $g = (g_1, g_2, \dots, g_n)$ is the least common multiple of the orders of the g_i if each of the elements g_i has finite order, and otherwise g has infinite order.*

PROOF. We have

$$g^m = (g_1^m, g_2^m, \dots, g_n^m),$$

and this is the identity if and only if m is a multiple of the orders of each of the g_i, so infinite if any one of them is infinite, and otherwise a multiple of the least common multiple. □

EXAMPLE 2.5.12. Let $G = \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$. Then every element of G has order dividing $\operatorname{lcm}(2,3,4,4) = 12$.

The latter example illustrates a more general phenomenon.

DEFINITION 2.5.13. The *exponent* of a group G is the smallest integer $n \ge 1$ such that $g^n = e$ for all $g \in G$, if it exists. Otherwise, it is infinite.

COROLLARY 2.5.14. *If G_i has exponent n_i for each $1 \le i \le n$, then the exponent of G is the least common multiple of the n_i.*

We mention the following result, the proof of which we leave to the reader.

PROPOSITION 2.5.15. *Suppose that $\{G_i \mid i \in I\}$ is a collection of groups and, for each $i \in I$, we are given $H_i \leqslant G_i$. Then we have*

$$\prod_{i \in I} H_i \leqslant \prod_{i \in I} G_i.$$

Note, however, that not all subgroups of a direct product are direct products of subgroups.

EXAMPLE 2.5.16. There are 5 subgroups of the *Klein four group* $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$:

$$\{0\}, \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}, \langle(1,0)\rangle, \langle(0,1)\rangle, \text{ and} \langle(1,1)\rangle.$$

The first four sit inside $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ as direct products of subgroups in the two individual coordinates, while the final subgroup does not.

Finally, we note the following interesting fact.

THEOREM 2.5.17. *Let m and n be relatively prime positive integers. Then the natural map*

$$\theta_{mn} \colon \mathbb{Z}/mn\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$$

induced by $a \mapsto (a,a)$ is an isomorphism. On the other hand, if m and n are not relatively prime, then $\mathbb{Z}/mn\mathbb{Z}$ and $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ are not isomorphic.

PROOF. Suppose that m and n are relatively prime. Note that

$$\theta_{mn}(a+b) = (a+b, a+b) = (a,a) + (b,b) = \theta_{mn}(a) + \theta_{mn}(b),$$

so θ_{mn} preserves the operation. If $(a,a) = (b,b)$ in $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$, then m and n both divide $a - b$, so mn does, as they are relatively prime. Therefore, θ_{mn} is injective. Since both groups have the same order mn, it is surjective as well.

If m and n are not relatively prime, then their least common multiple is

$$\mathrm{lcm}(m,n) = \frac{mn}{\gcd(m,n)} < mn.$$

Corollary 2.5.14 then implies that the exponent of $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ is less than mn, the exponent of $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$. As the exponent of a group is preserved by an isomorphism, the two groups in question cannot be isomorphic. □

The following equivalent corollary is known as the Chinese remainder theorem (CRT).

COROLLARY 2.5.18 (Chinese Remainder Theorem). *Let $k \geq 2$ and $m_1, \ldots, m_k$ be mutually relatively prime positive integers, which is to say that every pair of them is relatively prime. For any $b_1, \ldots, b_k \in \mathbb{Z}$, there exists an integer a, unique up to congruence modulo $m_1 m_2 \cdots m_k$, such that $a \equiv b_i \bmod m_i$ for each $1 \leq i \leq k$.*

PROOF. The existence in the case $k = 2$ is equivalent to the surjectivity of $\theta_{m_1 m_2}$ in Theorem 2.5.17, while the uniqueness is its injectivity. The case of general k follows by an easy induction on k. □

REMARK 2.5.19. We can give an explicit recipe for the construction of solutions of congruences modulo relatively prime integers (in the case of two congruences, and then by recursion). The construction is contained in the following direct proof that the map θ_{mn} in Theorem 2.5.17 is surjective:

Suppose that $b \in \mathbb{Z}/m\mathbb{Z}$ and $c \in \mathbb{Z}/n\mathbb{Z}$. Let $x, y \in \mathbb{Z}$ be such that $mx + ny \equiv 1 \bmod mn$, which we can find since $\gcd(m,n) = 1$. Then x is inverse to m in $\mathbb{Z}/n\mathbb{Z}$, and y is inverse to n in $\mathbb{Z}/m\mathbb{Z}$. Therefore, we have that

$$\theta_{mn}(cmx + bny) = (b,c) \in \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}.$$

EXAMPLE 2.5.20. Suppose we want to find $a \in \mathbb{Z}$ with

$$a \equiv 2 \bmod 7 \quad \text{and} \quad a \equiv 1 \bmod 5.$$

We note that 3 is an inverse of 7 modulo 5, and it is also an inverse of 5 modulo 7. So, in the proof of surjectivity in Theorem 2.5.17, we have $m = 7$, $n = 5$, $b = 2$, $c = 1$, $x = 3$, and $y = 3$, so

$$cmx + bny = 1 \cdot 7 \cdot 3 + 2 \cdot 5 \cdot 3 = 51 \equiv 16 \bmod 35.$$

Therefore $a = 16$ is the unique integer satisfying the two congruences. Moreover, note that x and y are independent of b and c, so we can use these x and y in solving any two congruences modulo 7 and 5.

EXAMPLE 2.5.21. We can use Theorem 2.5.17 to find isomorphisms between direct products of cyclic groups. For instance, using this and Remarks 2.5.4 and 2.5.5, we have that

$$\mathbb{Z}/12\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z} \cong \mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z} \cong \mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/24\mathbb{Z}.$$

2.6. Groups of isometries

DEFINITION 2.6.1. Let $n \geq 1$. An *isometry* of $\mathbb{R}^n$ is a bijection $f\colon \mathbb{R}^n \to \mathbb{R}^n$ that *preserves distances*, which is to say, for every $x, y \in \mathbb{R}^n$, we have

$$|f(x) - f(y)| = |x - y|,$$

where the absolute value denotes the distance from 0 in $\mathbb{R}^n$.

Since the property of preserving distances is preserved by composition, the following lemma is easily seen.

LEMMA 2.6.2. *The set $\mathscr{I}_n$ of isometries of $\mathbb{R}^n$ forms a group under composition.*

PROPOSITION 2.6.3. *If $f \in \mathscr{I}_1$, then there exists $a \in \mathbb{R}$ such that $f(x) = a + x$ or $f(x) = a - x$.*

REMARK 2.6.4. The function $f(x) = a + x$ is known as a *translation*, while the function $f(x) = a - x$ is known as a *reflection* (about the point $x = a$).

The group of isometries of $\mathbb{R}^2$ is a much more complicated group. We state, without proof, the following theorem.

THEOREM 2.6.5. *Every isometry f of $\mathbb{R}^2$ has one of the following four forms.*

i. f is a translation: there exists $a \in \mathbb{R}^2$ such that $f(x) = x + a$.

ii. f is a reflection: there exists a line L in $\mathbb{R}^2$ such that $f(x)$ is the reflection of x across L.

iii. f is a rotation: there exist $a \in \mathbb{R}^2$ and $\theta \in [0, 2\pi)$ such that f is given by counterclockwise rotation by θ radians about the center a.

iv. f is a glide reflection: f is the composition of a reflection and followed by a translation by a nonzero distance in a direction parallel to the line of reflection.

REMARK 2.6.6. Isometries of $\mathbb{R}^2$ are either *orientation-preserving* (i.e., the translations and the rotations) or *orientation-reversing* (i.e., the reflections and glide-reflections). An isometry that preserves orientation will map the letter “S” drawn in the plane to another letter than looks like an “S”, while an orientation-reversing isometry will map it to a backwards “S”. The composition of two orientation-preserving or two orientation-reversing isometries is orientation-preserving, while the composition of an orientation-preserving isometry with an orientation-reversing isometry (in either order) is orientation-reversing.

DEFINITION 2.6.7. Suppose that $X \subseteq \mathbb{R}^n$. We say that $f \in \mathscr{I}_n$ is a *symmetry* of X if for every $x \in \mathbb{R}^n$ one has $f(x) \in X$ if and only if $x \in X$.

The condition of f being a symmetry of X insures that the restriction of f to a map from X to X is a bijection.

DEFINITION 2.6.8. For $n \geq 3$, the *dihedral group D_n* is the group of symmetries of a regular n-gon, which we can take to be inscribed about the unit circle around the origin of $\mathbb{R}^2$ with a vertex at $(1, 0)$.

Note that a different choice of regular n-gon in $\mathbb{R}^2$ simply leads to an isomorphic group.

PROPOSITION 2.6.9. *For $n \geq 3$, the dihedral group D_n is a group of order $2n$, consisting of n rotations about the origin by multiples of $\frac{2\pi}{n}$ radians and n reflections. In the case that n is odd, these reflections are through lines through a vertex and a midpoint of the opposite side. In the case that n is even, $\frac{n}{2}$ of these reflections are through two opposite vertices and the other $\frac{n}{2}$ of them are through midpoints of two opposite sides.*

PROOF. The above-described rotations and reflections are all easily seen to be symmetries of the regular n-gon in question. We must see that these are the only ones. Any nonzero translation or glide reflection moves the origin of $\mathbb{R}^2$, and therefore moves the center of the polygon, hence gannot be a symmetry. Any rotation must be about the origin, or it too will move the center, and any rotation about the origin must take a vertex to a vertex, hence be by an angle that is a multiple of $2\pi/n$. Any reflection must for the same reason be a reflection across a line through the origin. If the line determining such a reflection does not cross a midpoint or vertex, then it will move the closest vertex on either side it passes through to a point which is less than the distance of a side of the polygon away, hence not to another vertex. Therefore, it must pass through the origin and either a midpoint or a vertex, and all such lines of reflection are described in the statement of the proposition. □

PROPOSITION 2.6.10. *Let $r \in D_n$ be counterclockwise rotation about the origin by $\frac{2\pi}{n}$ radians, and let $s \in D_n$ be the reflection across the x-axis. These two elements satisfy $r^n = 1$, $s^2 = 1$, and*

$$sr = r^{-1}s,$$

and every element in D_n may be written uniquely in the form $r^j s^k$ with $0 \leq j \leq n-1$ and $0 \leq k \leq 1$.

PROOF. The proposition boils down to the assertions that the r^j are all of the rotations and the $r^j s$ are all of the reflections. Since r^j is exactly rotation counterclockwise by $\frac{2\pi j}{n}$ radians, the first of these assertions holds. It is easy to see that if we first rotate across the x-axis and then rotate counterclockwise by $\frac{2\pi j}{n}$ radians, it is the same as reflecting across the line that is $\frac{\pi j}{n}$ radians counterclockwise from the x-axis. These lines pass alternately through vertices and midpoints for even and odd j, respectively, and hence are all of the reflections. □

COROLLARY 2.6.11. *We have $D_n = \langle r, s \rangle$, for r and s as in Proposition 2.6.10.*

There are many other interesting objects of which one can consider the symmetries, even in the plane. A pattern of finite, nonzero width and height that is repeated over an over infinitely in one direction has a symmetry group that is known as a “frieze group”, while a pattern of finite, nonzero width and height that is repeated over and over in two non-parallel directions is known as a “wallpaper group”.

2.7. Symmetric groups

Let n be a positive integer. In this section, we study the symmetric group S_n. Recall that an element of S_n is a bijection $\sigma \colon X_n \to X_n$, where X_n is the set $\{1, 2, \ldots, n\}$. It is common to denote the element σ of S_n by

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix}.$$

EXAMPLE 2.7.1. The permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 2 & 1 & 4 \end{pmatrix}$$

is the element of S_5 with values $\sigma(1) = 3$, $\sigma(2) = 5$, $\sigma(3) = 2$, $\sigma(4) = 1$, and $\sigma(5) = 4$.

This notation for permutations is amenable to composition.

EXAMPLE 2.7.2. Let σ be as in Example 2.7.1, and let

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 3 & 5 \end{pmatrix}.$$

To compute $\sigma\tau$, we write a three-by-three matrix with the top two rows given by the notation for τ and the next row determined by where σ takes the elements 1 through 5, i.e., we put $\sigma(i)$ below i for each i in the second row. This reads

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 3 & 5 \\ 5 & 3 & 1 & 2 & 4 \end{pmatrix}.$$

The first and third rows of the latter matrix then yield $\sigma\tau$:

$$\sigma\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 3 & 1 & 2 & 4 \end{pmatrix}.$$

Taking inverses is even easier: one merely switches the two rows.

EXAMPLE 2.7.3. Let σ be as in Example 2.7.1. Switching its two rows, we obtain

$$\begin{pmatrix} 3 & 5 & 2 & 1 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix},$$

and reordering the top row in the order 1 through 5, while preserving the columns by reordering the bottom row in the same fashion, we obtain

$$\sigma^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 1 & 5 & 2 \end{pmatrix},$$

DEFINITION 2.7.4. Let $\sigma \in S_n$. The *orbit* of $x \in X_n$ under σ is

$$O_\sigma(x) = \{\sigma^j(x) \mid j \in \mathbb{Z}\}.$$

EXAMPLE 2.7.5. Let σ be as in Example 2.7.1 and τ be as in Example 2.7.2. Then

$$O_\sigma(1) = \{1,2,3,4,5\},$$

while

$$O_\tau(1) = \{1,2\}, \quad O_\tau(3) = \{3,4\}, \quad O_\tau(5) = \{5\}.$$

PROPOSITION 2.7.6. *The relation $\sim$ on X_n given by $x \sim y$ if and only if $y = \sigma^i(x)$ for some $i \in \mathbb{Z}$ is an equivalence relation.*

PROOF. For $x \in X$, we have $x = \sigma^0(x)$, so $x \sim x$. For $x, y \in X$ with $x \sim y$, there exists $i \in \mathbb{Z}$ with $y = \sigma^i(x)$, and then $x = \sigma^{-i}(y)$, so $y \sim x$. For $x, y, z \in X$ with $x \sim y$ and $y \sim z$, we have $i, j \in \mathbb{Z}$ such that $y = \sigma^i(x)$ and $z = \sigma^j(y)$. We then have

$$z = \sigma^j(y) = \sigma^j(\sigma^i(x)) = \sigma^{i+j}(x),$$

so $x \sim z$. □

By definition, $O_\sigma(x)$ is the equivalence class of x under the equivalence relation defined in Propositiion 2.7.6.

DEFINITION 2.7.7.

a. For $k \geq 2$, a *k-cycle* in S_n is an element of S_n that has one orbit with k elements, and for which all the other orbits have only one element each.

b. A *cycle* is a permutation that is a k-cycle for some $k \geq 2$.

c. The *length* of a cycle σ is the integer $k \geq 2$ such that σ is a k-cycle.

EXAMPLE 2.7.8. The element σ of Example 2.7.1 is a 5-cycle, but τ as in Example 2.7.2 is not a cycle.

We have another notation for permutations, which depends on their orbit decomposition. We begin with the case of a cycle.

NOTATION 2.7.9. Suppose that σ is a k-cycle, and let x be an element in its largest orbit. We use

$$(x\ \sigma(x)\ \cdots\ \sigma^{k-1}(x))$$

to denote the element σ.

REMARK 2.7.10. When a group G is viewed as a subgroup of a symmetric group (i.e., as consisting of permutations of some set), any non-identity element of G is called a *nontrivial element* of G, and the identity element is called *trivial*, or the *trivial element*.

There are k different ways to write a k-cycle in the form of Definition 2.7.9.

EXAMPLE 2.7.11. The 5-cycle σ of Example 2.7.1 is equal to

$$(1\ 3\ 2\ 5\ 4) = (3\ 2\ 5\ 4\ 1) = (2\ 5\ 4\ 1\ 3) = (5\ 4\ 1\ 3\ 2) = (4\ 1\ 3\ 2\ 5).$$

EXAMPLE 2.7.12. Every nontrivial element of S_3 is a cycle: these elements are (12), (13), (23), (123), and (132).

However, not every nontrivial element of S_4 is a cycle, as a permutation in S_4 can have two orbits of order 2.

DEFINITION 2.7.13. We say that two cycles σ and τ in S_n are *disjoint* if the largest orbit of σ has empty intersection with the largest orbit of τ.

We prove the following lemma.

LEMMA 2.7.14. *Any two disjoint cycles commute.*

PROOF. Let $\sigma, \tau \in S_n$ be disjoint cycles. Let O_σ and O_τ denote their largest orbits. If $x \in O_\sigma$, then $x, \sigma(x) \notin O_\tau$, so

$$\sigma\tau(x) = \sigma(x) = \tau\sigma(x).$$

Similarly, σ and τ commutate on elements of O_τ. Finally, if x lies in the complement of $O_\sigma \cup O_\tau$, then $\sigma\tau(x) = x = \tau\sigma(x)$. So τ and σ commute. □

REMARK 2.7.15. We see from the proof of Lemma 2.7.14 that in any product τ of disjoint cycles, an element $x \in X_n$ will be fixed by all but at most one of the cycles, and if there is such a cycle, the value $\tau(x)$ will equal the value of that cycle on x (and otherwise $\tau(x) = x$).

We next see that the cycles generate S_n.

PROPOSITION 2.7.16. *Every nontrivial permutation in S_n may be written as a product of disjoint cycles in a unique way, up to the order of the cycles.*

PROOF. Let $\sigma \in S_n$, and suppose that σ has m orbits of order greater than 1. Choose representatives x_i of each of these orbits for $1 \le i \le m$, and set $k_i = |O_\sigma(x_i)|$. We then consider the product

$$\tau = (x_1\ \sigma(x_1)\ \cdots\ \sigma(x_1)^{k_1-1})(x_2\ \sigma(x_2)\ \cdots\ \sigma(x_2)^{k_2-1})\cdots(x_m\ \sigma(x_m)\ \cdots\ \sigma(x_m)^{k_m-1})$$

of disjoint cycles, and we claim that $\tau = \sigma$. For $x \in X_n$, we have either $x = \sigma^j(x_i)$ for some $1 \le i \le m$ and $0 \le j \le k_i - 1$, or x lies in an orbit of order 1. In the former case, we have $\tau(x) = \sigma^{j+1}(x_i) = \sigma(x)$. In the latter, we have $\tau(x) = x = \sigma(x)$ as well. Hence, we see that σ may be written as a product of disjoint cycles.

We leave uniqueness primarily to the reader. We merely note that, first, the elements appearing in the individual cycles above are the elements in the orbits and so must be in any such decomposition of σ. Given that, the individual cycles are forced to be as above by the values of σ. □

EXAMPLE 2.7.17. Consider the permutation

$$\lambda = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 6 & 4 & 3 & 5 & 1 \end{pmatrix}.$$

Then $\lambda(1) = 2$, $\lambda(2) = 6$, $\lambda(6) = 1$, so one of the cycles in the decomposition of λ is $(1\ 2\ 6)$. Also, $\lambda(3) = 4$ and $\lambda(4) = 3$, so another is $(3\ 4)$. On the other hand, $\lambda(5) = 5$, so 5 is not moved (or is "fixed") by λ. We therefore have

$$\lambda = (1\ 2\ 6)(3\ 4).$$

Proposition 2.7.16 has the following interesting application to orders of elements.

PROPOSITION 2.7.18. *The order of an element of S_n is the least common multiple of the orders of the disjoint cycles of which it is a product.*

PROOF. Suppose that $\sigma \in S_n$ decomposes as a product of m disjoint cycles $\tau_1, \tau_2, \ldots, \tau_m$ of length $k_1, k_2, \ldots, k_m$. Then Lemma 2.7.14 implies that

$$\sigma^i = \tau_1^i \cdots \tau_m^i$$

for every $i \in \mathbb{Z}$, and by disjointness, the only way for $\sigma^i = e$ to occur is if $\tau_j^i = e$ for all $1 \leq j \leq m$. But τ_j has order k_j, so this will happen if and only if i is a multiple of each k_j, and therefore of the least common multiple of the k_j. Hence the order of σ is this least common multiple. □

EXAMPLE 2.7.19. In S_7, the element

$$\sigma = (1\ 4\ 2\ 7)(3\ 6) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 7 & 6 & 2 & 5 & 3 & 1 \end{pmatrix}$$

has order 4. Note that

$$\sigma^2 = (1\ 2)(4\ 7) \quad \text{and} \quad \sigma^3 = (1\ 7\ 2\ 4)(3\ 6).$$

EXAMPLE 2.7.20. The exponent of a finite group is the least common multiple of the orders of its elements. Since the order of an element is the least common multiple of the orders of its cycles, which have orders $2, \ldots, n$, the exponent of S_n is $\mathrm{lcm}(1, 2, \ldots, n)$. E.g., the exponent of S_7 is $3 \cdot 4 \cdot 5 \cdot 7 = 420$.

In fact, G has a smaller generating set than the cycles, which is to say the set of transpositions.

DEFINITION 2.7.21. A *transposition* in S_n is a 2-cycle.

PROPOSITION 2.7.22. *Every element of S_n is a product of transpositions.*

PROOF. As every permutation is a product of cycles, we need only show that every cycle is a product of transpositions. In fact,

$$(x_1\ x_2\ \ldots\ x_k) = (x_1\ x_2)(x_2\ x_3) \cdots (x_{k-1}\ x_k),$$

as is easily checked. □

REMARK 2.7.23. In fact, the symmetric group S_n is generated by transpositions of the form $(x\ x+1)$ with $1 \leq x \leq n-1$. That is, the previous proposition tells us that

$$(x\ x+1\ \cdots y) = (x\ x+1)(x+1\ x+2) \cdots (y-1\ y)$$

for any $1 \leq x < y \leq n$, and then for such x and y we have

$$(x\ y) = (y-1\ y) \cdot (y-2\ y-1\ y) \cdots (x\ x+1 \cdots y),$$

so all transpositions are contained in the subgroup generated by transpositions of the form $(x\ x+1)$, which again by Proposition 2.7.22 is all that we need.

2.8. Homomorphisms

In order to compare groups, it is useful to consider a generalization of the concept of isomorphism that actually has a simpler definition, as the condition of bijectivity is removed.

DEFINITION 2.8.1. Let G and G' be groups. A *homomorphism* ϕ from G to G' is a function

$$\phi \colon G \to G'$$

such that

$$\phi(ab) = \phi(a)\phi(b)$$

for all $a, b \in G$.

EXAMPLES 2.8.2.

a. Let $\psi_n\colon \mathbb{Z} \to \mathbb{Z}$ be the multiplication-by-n map, defined by $\psi_n(a) = na$ for all $a \in \mathbb{Z}$. Then ψ_n is a homomorphism since

$$\psi_n(a+b) = n(a+b) = na+nb = \psi_n(a)+\psi_n(b).$$

b. The reduction map $\phi_n\colon \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ defined by $\phi_n(a) = a+n\mathbb{Z}$ is a surjective homomorphism.

c. The determinant map

$$\det\colon \mathrm{GL}_n(\mathbb{R}) \to \mathbb{R}^\times$$

satisfies

$$\det(AB) = \det(A)\det(B)$$

for all $A,B \in \mathrm{GL}_n(\mathbb{R})$, so is a (surjective) homomorphism.

d. For $m < n$, we have an (injective) homomorphism $\iota\colon S_m \to S_n$ that takes a permutation σ of $X_m = \{1,2,\ldots,m\}$ to the permutation $\tau \in S_n$ that satisfies $\tau(i) = \sigma(i)$ for $1 \le i \le m$ and $\tau(j) = j$ for $m < j \le n$.

e. For $A \in M_{rs}(\mathbb{R})$, we define a left-multiplication-by-A map

$$\psi_A\colon M_{st}(\mathbb{R}) \to M_{rt}(\mathbb{R})$$

by $\psi_A(B) = AB$ for $B \in M_{st}(\mathbb{R})$. By distributivity of multiplication of matrices, this is a homomorphism. It need not in general be injective or surjective.

f. The set

$$C^1(\mathbb{R}) = \{f\colon \mathbb{R} \to \mathbb{R} \mid f \text{ is everywhere differentiable}\}$$

forms a group under addition. In fact, it is a subgroup of $\mathrm{Maps}(\mathbb{R},\mathbb{R})$. The derivative map

$$\partial\colon C^1(\mathbb{R}) \to \mathrm{Maps}(\mathbb{R},\mathbb{R}), \qquad \partial(f) = f'$$

is a homomorphism.

Here are several standard homomorphisms between groups.

DEFINITION 2.8.3. Let G and G' be groups with identity elements e and e', respectively.

a. The *trivial homomorphism* $\phi\colon G \to G'$ is given by $\phi(g) = e'$, the identity of G', for all $g \in G$.

b. The *identity homomorphism* $\mathrm{id}_G\colon G \to G$ on any group G, given by $\mathrm{id}_G(g) = g$ for all $g \in G$.

c. For $H \leqslant G$, we the *inclusion map* $\iota_H\colon H \to G$ with $\iota_H(h) = h$ for all $h \in H$.

The following easily-proven lemma is useful to know.

LEMMA 2.8.4. *Let G, G', and G'' be groups, and let $\phi\colon G \to G'$ and $\psi\colon G' \to G''$ be homomorphisms. Then*

$$\psi \circ \phi\colon G \to G''$$

is also a homomorphism.

PROOF. For $a, b \in G$, we have

$$\psi \circ \phi(ab) = \psi(\phi(a)\phi(b)) = \psi(\phi(a))\psi(\phi(b)) = \psi \circ \phi(a) \cdot \psi \circ \phi(b).$$

□

In the following, G and G' will be groups, and we will use e and e' to denote their respective identity elements.

LEMMA 2.8.5. *Let $\phi\colon G \to G'$ be a homomorphism. Then $\phi(e) = e'$, and $\phi(g^{-1}) = \phi(g)^{-1}$ for all $g \in G$.*

PROOF. We have $\phi(e) = \phi(e)\phi(e)$ by the defining property of a homomorphism, and the cancellation theorem then implies that $\phi(e) = e'$. Moreover,

$$\phi(g) \cdot \phi(g^{-1}) = \phi(e) = e',$$

again by the homomorphism property, and uniqueness of right inverses in a group then implies that $\phi(g^{-1}) = \phi(g)^{-1}$. The last statement then follows easily from these and the homomorphism property. □

We have the following easy consequence.

LEMMA 2.8.6. *Let $\phi\colon G \to G'$ be a homomorphism. Then*

$$\phi(g_1^{r_1} g_2^{r_2} \cdots g_k^{r_k}) = \phi(g_1)^{r_1}\phi(g_2)^{r_2} \cdots \phi(g_k)^{r_k}$$

for any $k \geq 1$, $g_1, g_2, \dots, g_k \in G$, and $r_1, r_2, \dots, r_k \in \mathbb{Z}$.

PROOF. For $g \in G$ and $r \in \mathbb{Z}$, we have $g^r = (g^{-1})^{-r}$ and

$$\phi(g)^r = \phi(g^{-1})^{-r},$$

so it suffices to assume that each r_i is nonnegative in the theorem. But then, by writing out the powers as products, the result amounts simply to proving the result when each $r_i = 1$. On the other hand,

$$\phi(g_1 g_2 \cdots g_k) = \phi(g_1)\phi(g_2 \cdots g_k) = \cdots = \phi(g_1)\phi(g_2) \cdots \phi(g_k)$$

by iterative use of the defining property of a homomorphism. □

DEFINITION 2.8.7. Let $\phi\colon G \to G'$ be a homomorphism.

a. The *kernel* of ϕ is the subset of G that is

$$\ker \phi = \{g \in G \mid \phi(g) = e'\}.$$

b. The *image* of ϕ is the subset of G' that is

$$\operatorname{im} \phi = \{\phi(g) \mid g \in G\}.$$

PROPOSITION 2.8.8. *Let $\phi\colon G \to G'$ be a homomorphism. Then* $\ker \phi$ *is a subgroup of G and* $\operatorname{im} \phi$ *is a subgroup of G'.*

PROOF. Since $\phi(e) = e'$, we have $e \in \ker\phi$. Moreover, if $a, b \in \ker\phi$ then

$$\phi(ab) = \phi(a)\phi(b) = e' \cdot e' = e',$$

so $ab \in \ker\phi$, and if $a \in \ker\phi$ then

$$\phi(a^{-1}) = \phi(a)^{-1} = (e')^{-1} = e',$$

so $a^{-1} \in \ker\phi$. It follows that $\ker\phi \leqslant G$.

Next, note that $e' = \phi(e)$, so $e' \in G$. Also, if $\phi(a), \phi(b) \in \operatorname{im}\phi$ for some $a, b \in G$, then

$$\phi(a)\phi(b) = \phi(ab) \in \operatorname{im}\phi$$

and

$$\phi(a)^{-1} = \phi(a^{-1}) \in \operatorname{im}\phi.$$

Hence, we have that $\operatorname{im}\phi \leqslant G'$. □

Clearly, a homomorphism $\phi\colon G \to G'$ is surjective if and only if $\operatorname{im}\phi = G'$. On the other hand, we have the following less obvious criterion for injectivity of ϕ in terms of its kernel.

PROPOSITION 2.8.9. *A homomorphism $\phi\colon G \to G'$ is injective if and only if* $\ker\phi = \{e\}$.

PROOF. If ϕ is injective and $a \in \ker\phi$, then $\phi(a) = e' = \phi(e)$, so $a = e$ by injectivity of ϕ. On the other hand, if $\ker\phi$ is trivial and $\phi(a) = \phi(b)$ for some $a, b \in G$, then

$$\phi(ab^{-1}) = \phi(a)\phi(b)^{-1} = e',$$

so $ab^{-1} = e$, and therefore $a = b$. □

EXAMPLES 2.8.10.

a. The multiplication-by-n map ψ_n is injective, as $\psi_n(a) = na = 0$ if and only if $a = 0$. Its image is $n\mathbb{Z}$.

b. The inclusion map $\iota_H\colon H \to G$ of a subgroup H in a group G is obviously injective, and its image is H.

c. The reduction map $\phi_n\colon \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ is surjective, and its kernel is $n\mathbb{Z}$.

d. The determinant map $\det\colon \mathrm{GL}_n(\mathbb{R}) \to \mathbb{R}^\times$ is surjective with kernel $\mathrm{SL}_n(\mathbb{R})$.

e. The derivative map $\partial\colon C^1(\mathbb{R}) \to \mathrm{Maps}(\mathbb{R}, \mathbb{R})$ has kernel equal to the subgroup of constant functions. Its image is difficult to describe explicitly, but it is not surjective.

We can also speak of the image of a subgroup under a homomorphism.

DEFINITION 2.8.11. Let H be a subgroup of G. Then *image* of H under a homomorphism $\phi\colon G \to G'$ is

$$\phi(H) = \{\phi(h) \mid h \in H\}.$$

REMARK 2.8.12. The set $\phi(H)$ is a subgroup of G', as it is the image of composition $\phi \circ \iota_H$ of the inclusion map $\iota_H\colon H \to G$ with ϕ.

DEFINITION 2.8.13. The *restriction of a homomorphism* $\phi\colon G \to G'$ to $H \leqslant G$ is the homomorphism

$$\phi|_H\colon H \to G'$$

that is the composition $\phi \circ \iota_H$, where ι_H is the inclusion map. In other words, $\phi|_H(h) = \phi(h)$ for all $h \in H$.

We can also speak of the inverse image of a subgroup under a homomorphism.

DEFINITION 2.8.14. Let $\phi\colon G \to G'$ be a homomorphism. Let $H' \leqslant G'$. The *inverse image* of H' under ϕ is

$$\phi^{-1}(H') = \{h \in H \mid \phi(h) \in H'\}.$$

PROPOSITION 2.8.15. *Let $\phi\colon G \to G'$ be a homomorphism, and let $H' \leqslant G'$. Then $\phi^{-1}(H')$ is a subgroup of G.*

PROOF. Note that $\phi(e) = e' \in H'$, so $e \in \phi^{-1}(H')$. Also, if $a, b \in \phi^{-1}(H')$, then $\phi(ab^{-1}) = \phi(a)\phi(b)^{-1} \in H'$, since H' is a group, so $ab^{-1} \in \phi^{-1}(H')$. In other words, $\phi^{-1}(H')$ is closed under multiplcation and inverses, so is a subgroup of G. □

EXAMPLE 2.8.16. Consider the multiplication-by-n map $\psi_n\colon \mathbb{Z} \to \mathbb{Z}$. We have

$$\psi_n^{-1}(m\mathbb{Z}) = \{a \in \mathbb{Z} \mid na \in m\mathbb{Z}\} = \frac{m}{\gcd(n,m)}\mathbb{Z}.$$

A homomorphism is completely determined by its values on a generating set.

PROPOSITION 2.8.17. *Let S be a generating set of G, and let $\phi, \psi\colon G \to G'$ be homomorphisms. Suppose that $\phi(s) = \psi(s)$ for all $s \in S$. Then $\phi = \psi$.*

PROOF. Since S generates G, every element of G has the form

$$s_1^{r_1} s_2^{r_2} \cdots s_k^{r_k}$$

for some $k \geq 0$, $s_1, s_2, \ldots, s_k \in G$, and $r_1, r_2, \ldots, r_k \in \mathbb{Z}$. We have

$$\phi(s_1^{r_1} s_2^{r_2} \cdots s_k^{r_k}) = \phi(s_1)^{r_1}\phi(s_2)^{r_2}\cdots\phi(s_k)^{r_k} = \psi(s_1)^{r_1}\psi(s_2)^{r_2}\cdots\psi(s_k)^{r_k} = \psi(s_1^{r_1} s_2^{r_2} \cdots s_k^{r_k}),$$

as desired. □

This is a very useful property for checking whether or not two homomorphisms are equal. On the other hand, one might be tempted to try to use it to specify a homomorphism by setting its values on a generating set arbitrarily. This in general does not work. For instance, the only homomorphism $\phi\colon \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}$ is the trivial homomorphism, since one must have $n\phi(1) = \phi(n) = 0$. That is, one can't simply take $\phi(1)$ to be an arbitrary value of $\mathbb{Z}$.

We leave it to the reader to check the following easy assertion, which in particular explains the problem just described.

LEMMA 2.8.18. *Suppose that $g \in G$ has finite order and $\phi\colon G \to G'$ is a homomorphism. Then the order of $\phi(g)$ divides the order of g.*

2.9. The alternating group

In this section, we study a certain subgroup of S_n, known as the alternating group. Let us begin with its definition.

DEFINITION 2.9.1. The *alternating group* A_n is the subgroup of S_n consisting of permutations that can be written as a product of an even number of transpositions.

EXAMPLE 2.9.2. We have $(1\ 2\ 3) \in A_3$, as $(1\ 2\ 3) = (1\ 2)(2\ 3)$. Even more obviously, we have $(1\ 2)(3\ 4) \in A_4$.

The assertion that A_n is a subgroup of S_n contained in Definition 2.9.1 is easy to verify, and we leave it to the reader. What is not so immediate is that not every element in S_n can be written as a product of an even number of transpositions, which is to say that $A_n \neq S_n$. For this reason and others, we give an alternate characterization of S_n.

DEFINITION 2.9.3. A *permutation matrix* in $\mathrm{GL}_n(\mathbb{R})$ is a matrix for which the entries are all zero aside from one entry in each row and each column, which is 1.

PROPOSITION 2.9.4. *The set P_n of permutation matrices forms a subgroup of* $\mathrm{GL}_n(\mathbb{R})$ *that is isomorphic to S_n. Explicitly, define*

$$\kappa \colon S_n \to \mathrm{GL}_n(\mathbb{R})$$

by taking $\kappa(\sigma)$ to be the matrix with entry

$$\kappa(\sigma)_{ij} = \begin{cases} 1 & \textit{if } \sigma(j) = i \\ 0 & \textit{if } \sigma(j) \neq i. \end{cases}$$

in the ith row and jth column. Then κ is injective with image P_n.

PROOF. Let $\sigma \in S_n$. We first remark that $\kappa(\sigma)$ is a permutation matrix: since σ is a function, each j is taken to exactly one i, so each column has exactly one 1, and since σ is a bijection, each i comes from exactly one j, so each row has exactly one 1. Moreover, we have

$$(\kappa(\sigma)\kappa(\tau))_{ik} = \sum_{j=1}^{n} \kappa(\sigma)_{ij}\kappa(\tau)_{jk} = \kappa(\sigma)_{i\tau(k)}\kappa(\tau)_{\tau(k)k} = \begin{cases} 1 & \text{if } \sigma(\tau(k)) = i \\ 0 & \text{if } \sigma(\tau(k)) \neq i, \end{cases}$$

and the latter term is exactly $\kappa(\sigma\tau)_{ik}$. Therefore, κ is a homomorphism. It is also clearly one-to-one, since $\kappa(\sigma)$ will have a non-diagonal entry that is nonzero if $\sigma \neq e$. Finally, for any $A = (a_{ij}) \in P_n$, we have $A = \kappa(\sigma)$, where $\sigma(j)$ is defined as the unique i such that $a_{ij} = 1$, so κ is onto. □

We next determine the image of A_n under the map κ of Proposition 2.9.4.

LEMMA 2.9.5. *The image of A_n under κ is equal to the subgroup of permutation matrices that have determinant* 1.

PROOF. Then $\kappa((a\ b))$, where $1 \le a < b \le n$, is exactly the identity matrix after one row operation, which is switching the ith and jth rows. As switching two rows changes the sign of a matrix, we have $\det\kappa((a\ b)) = -1$. As $\det\circ\kappa$ is a homomorphism, we have that the determinant of a product of a product of an even number of transpositions is 1, and the determinant of the product of an odd number of transpositions is -1. □

For the following definition, we note that the determinant of any permutation matrix is either 1 or -1.

DEFINITION 2.9.6.

a. We define the *sign function* on S_n by

$$\mathrm{sign} = \det \circ \kappa \colon S_n \to \{\pm 1\},$$

with κ as in Proposition 2.9.4. Its value on a permutation is the *sign* of the permutation.

b. We say that $\sigma \in S_n$ is *even* if $\mathrm{sign}(\sigma) = 1$ and *odd* if $\mathrm{sign}(\sigma) = -1$.

REMARK 2.9.7. By Lemma 2.9.5, we have that A_n is exactly the subgroup of even permutations in S_n.

EXAMPLE 2.9.8. Cycles of even length are odd, while cycles of odd length are even.

2.10. Cosets

DEFINITION 2.10.1. Let H be a subgroup of a group G, and let a be an element of G. The *left H-coset* of a is the subset of G that is

$$aH = \{ah \mid h \in H\}.$$

The *right H-coset* of a is the set

$$Ha = \{ha \mid h \in H\}.$$

REMARK 2.10.2. If G is abelian, then $aH = Ha$ for any $H \leqslant G$ and $a \in G$, so we may speak simply of cosets (as opposed to left and right cosets). If the operation on G is addition, we write $a + H$ for the H-coset of a.

EXAMPLE 2.10.3. The $2\mathbb{Z}$-coset $1 + 2\mathbb{Z}$ of 1 in $\mathbb{Z}$ is the set of odd integers.

EXAMPLE 2.10.4. Let $H = \langle (12) \rangle \leqslant S_3$. Then the left cosets of H are

$$\begin{gathered} H = (1\ 2)H = \{(1\ 2), e\} \\ (1\ 2\ 3)H = (1\ 3)H = \{(1\ 3), (1\ 2\ 3)\} \\ (1\ 3\ 2)H = (2\ 3)H = \{(2\ 3), (1\ 3\ 2)\}. \end{gathered}$$

The property of two cosets being equal provides an equivalence relation on a group G, as expressed in the following lemma, the proof of which follows directly from the definitions of left and right cosets.

LEMMA 2.10.5. *Let H be a subgroup of a group G, and let $a, b \in H$. The relation $a \sim_l b$ (resp., $a \sim_r b$) if and only if $aH = bH$ (resp., $Ha = Hb$) is an equivalence relation on G, and the equivalence class of $a \in H$ under this relation is aH (resp., Ha).*

COROLLARY 2.10.6. *If $H \leqslant G$, then G is the disjoint union of its distinct left (or right) H-cosets.*

We make the following remark.

LEMMA 2.10.7. *Let $H \leqslant G$, and let $a, b \in H$. Then $aH = bH$ if and only if $a^{-1}b \in H$, and $Ha = Hb$ if and only if $ab^{-1} \in H$.*

PROOF. Suppose $a^{-1}b \in H$, and set $h = a^{-1}b$. Then $b = ah$, so $bk = a(hk) \in aH$ for every $k \in H$, which implies $bH \subseteq aH$. Moreover, $bh^{-1} = a$, so $aH \subseteq bH$ as well. Conversely, if $aH = bH$, then there exists $h \in H$ such that $b = ah$, so $a^{-1}b \in H$. The case of right cosets is similar, noting that $ab^{-1} \in H$ if and only if $a = hb$ for some $h \in H$. □

NOTATION 2.10.8. For $H \leqslant G$, we let G/H denote the set of left cosets of H in G, and we let $H\backslash G$ be the set of right cosets of H in G.

EXAMPLE 2.10.9. One might notice that the set of left cosets of $n\mathbb{Z}$ in $\mathbb{Z}$ is given the notation $\mathbb{Z}/n\mathbb{Z}$ by Corollary 2.10.8, which could in theory lead to some confusion with the group $\mathbb{Z}/n\mathbb{Z}$. However, this is no coincidence. The cosets of $n\mathbb{Z}$ in $\mathbb{Z}$ are the $a + n\mathbb{Z}$ with $0 \leq a \leq n-1$, which are exactly the elements of the group $\mathbb{Z}/n\mathbb{Z}$. So, $\mathbb{Z}/n\mathbb{Z}$ as a group is just the set $\mathbb{Z}/n\mathbb{Z}$ (of cosets) with a particular binary operation.

DEFINITION 2.10.10. We refer to a set of representatives for the left (resp., right) H-cosets in G as a set of *left* (resp., *right*) *coset representatives*.

EXAMPLE 2.10.11. The elements $e, (1\ 2\ 3), (1\ 3\ 2)$ form a set of left coset representatives for $\langle (1\ 2) \rangle$ in S_3.

EXAMPLE 2.10.12. The coset $s\langle r \rangle = \langle r \rangle s$ consists of all reflections in D_n. The coset $r^k\langle s \rangle = \{r^k, r^k s\}$ consists of the counterclockwise rotation r^k about the origin by $2\pi k/n$ radians and the reflection $r^k s$ across the line through the origin at an angle $\pi k/n$ radians counterclockwise from the x-axis. Note that

$$\langle s \rangle r^k = \{r^k, r^{-k}s\},$$

and this is not $r^k\langle s \rangle$ unless $2k = n$.

PROPOSITION 2.10.13. *There is a canonical bijection*

$$\phi \colon G/H \to H\backslash G$$

given by $\phi(aH) = Ha^{-1}$ for $a \in G$.

PROOF. First, we check that the map ϕ is well-defined. For a subset A of G, let us use A^{-1} to denote

$$A^{-1} = \{a^{-1} \mid a \in A\}.$$

If $aH = bH$, then

$$Ha^{-1} = \{ha^{-1} \mid h \in H\} = \{(ah^{-1})^{-1} \mid h \in H\} = \{(ak)^{-1} \mid k \in H\} = (aH)^{-1}. \tag{2.10.1}$$

Since $aH = bH$, we have that $(aH)^{-1} = (bH)^{-1}$, which implies noting (2.10.1) for both a and b that

$$Ha^{-1} = (aH)^{-1} = (bH)^{-1} = Hb^{-1}.$$

Therefore, ϕ is well-defined.

Next, define $\psi\colon H\backslash G \to G/H$ by $\psi(Ha) = a^{-1}H$ for $a \in G$. This is also well-defined, as $Ha = Hb$ implies that

$$a^{-1}H = (Ha)^{-1} = (Hb)^{-1} = b^{-1}H,$$

and it is clearly inverse to ϕ, so ϕ is a bijection. □

EXAMPLE 2.10.14. Let $H = \langle(1\ 2)\rangle \leqslant S_3$. Then the bijection ϕ of Proposition 2.10.13 is given by $\phi(H) = H$,

$$\phi((1\ 2\ 3)H) = H(1\ 3\ 2), \quad \text{and} \quad \phi((1\ 3\ 2)H) = H(1\ 2\ 3).$$

DEFINITION 2.10.15. Let H be a subgroup of G.

a. If there are finitely many left cosets of H in G, then we say that H is of *finite index* in G, and otherwise H is of *infinite index*.

b. If H is of finite index in G, then we define the *index* $[G:H]$ of H in G to be the number of left cosets of H in G.

EXAMPLES 2.10.16.

a. The index $[S_3:H]$ of $H = \langle(12)\rangle$ in S_3 is 3.

b. The index $[\mathbb{Z}:n\mathbb{Z}]$ of $n\mathbb{Z}$ in $\mathbb{Z}$ is n.

c. The group $\mathbb{Z}$ is not of finite index in $\mathbb{Q}$.

d. We have $[D_n:\langle s\rangle] = n$ and $[D_n:\langle r\rangle] = 2$.

e. We have $[S_n:A_n] = 2$ for $n \geq 2$, and the nonidentity coset is the set of odd permutation which equals, e.g., $(1\ 2)A_n$.

EXAMPLE 2.10.17. For any group G, we have $[G:G] = 1$, and if G is finite, we have $[G:\langle e\rangle] = |G|$

REMARK 2.10.18. By Proposition 2.10.13, we could just as well have used right cosets instead of left cosets in the definition of the index.

THEOREM 2.10.19 (Lagrange's theorem). *Let H be a subgroup of a finite group G. Then we have*

$$|G| = [G:H]|H|.$$

In particular, the order of H divides the order of G.

PROOF. Since G is finite, so is H, and every coset aH is in bijection with H via the map $\theta\colon H \to aH$ with $\theta(h) = ah$. As G is the disjoint union of its left cosets, we have

$$|G| = \sum_{aH\in G/H} |aH| = \sum_{aH\in G/H} |H| = [G:H]|H|.$$

□

We can use Lagrange's theorem to determine the indices of subgroups when we know both the orders of the group and of the subgroup. Here is an example.

EXAMPLE 2.10.20. There is an injective homomorphism

$$\iota \colon S_{n-1} \to S_n$$

as in Example 2.8.2 that takes a permutation of the set $X_{n-1} \subset X_n$ to the permutation that has the same values on the elements of X_{n-1} and which fixes n. Using ι, we may identify S_{n-1} with the isomorphic subgroup $\iota(S_{n-1})$ of S_n consisting of elements $\sigma \in S_n$ with $\sigma(n) = n$. Under this identification, we have

$$[S_n : S_{n-1}] = \frac{n!}{(n-1)!} = n.$$

EXAMPLE 2.10.21. Since $[S_n : A_n] = 2$ for $n \geq 2$, we have $|A_n| = \frac{1}{2}n!$ for such n.

COROLLARY 2.10.22. *Let G be a finite group. Then the order of every element of G divides the order of G.*

EXAMPLE 2.10.23. We have already seen that the orders of the subgroups of $\mathbb{Z}/n\mathbb{Z}$ are exactly the positive divisors of n.

EXAMPLE 2.10.24. According to the corollary, every element of S_n should have order dividing $n!$. In fact, we already know from Example 2.7.20 that every element has order dividing the least common multiple of $1, 2, \ldots, n$, which clearly divides $n!$.

Finally, we mention the following interesting corollary of Lagrange's theorem.

COROLLARY 2.10.25. *Every group of prime order is cyclic.*

PROOF. Let G be a group of order a prime p. If $g \in G$ is not the identity, it must generate a nontrivial subgroup of G, which can only have order p by Lagrange's theorem, and therefore must be G. That is, $G = \langle g \rangle$, finishing the proof. □

We note that the index satisfies the following multiplicative property.

PROPOSITION 2.10.26. *Let H and K be subgroups of G with $K \leqslant H$. Then K has finite index in G if and only if H has finite index in G and K has finite index in H. Moreover, if K has finite index in G, then we have*

$$[G : K] = [G : H][H : K].$$

PROOF. Let S be a set of H-coset representatives in G and T be a set of K-coset representatives in H. Consider the set

$$U = \{st \mid s \in S, t \in T\}.$$

We claim that U is a set of left K-coset representatives in G. For this, note that if $g \in G$, we may choose $s \in S$ with $g = sh$ for some $h \in H$, and we may choose $t \in T$ with $h = tk$ for some $k \in K$. In other words, $g \in stK$, so $gK = stK$. This proves the claim.

Next, note that if $stK = s't'K$ with $s, s' \in S$ and $t, t' \in T$, then since $tK \subset H$ and $t'K \subset H$, we have $sH = s'H$, so $s = s'$. But then $stK = st'K$, so $tK = t'K$, and therefore $t = t'$. In other words, we have shown that the map $S \times T \to U$ given by $(s,t) \mapsto st$ is a bijection. Thus, S and T are finite if and only if U is, and if they are, then $|U| = |S||T|$, as desired. □

REMARK 2.10.27. Proposition 2.10.26 implies Lagrange's theorem by taking the subgroup K to be the trivial subgroup. That is, for a finite group G and subgroup H, we have

$$|G| = [G : \langle e\rangle] = [G : H][H : \langle e\rangle] = [G : H]|H|.$$

2.11. Conjugation

DEFINITION 2.11.1. Let G be a group.

a. Let $a, x \in G$. Then axa^{-1} is known as the *conjugate* of x by a.

b. We say that an element $x \in G$ is *conjugate* to an element $y \in G$ if there exists $a \in G$ with $y = axa^{-1}$.

REMARK 2.11.2. One might recall the related notion of similar matrices in $M_n(\mathbb{R})$.

LEMMA 2.11.3. *The relation $\sim$ on G given by $x \sim y$ if and only if x is conjugate to y is an equivalence relation on G.*

PROOF. We have $x = exe^{-1}$, so $x \sim x$. If $x \sim y$, then there exists $a \in G$ with $y = axa^{-1}$, which implies

$$x = a^{-1}ya = a^{-1}y(a^{-1})^{-1},$$

so $y \sim x$. Finally, if $x \sim y$ and $y \sim z$, then there exist $a,b \in G$ with $y = axa^{-1}$ and $z = byb^{-1}$, so

$$z = byb^{-1} = b(axa^{-1})b^{-1} = (ba)x(ba)^{-1},$$

and $z \sim x$. □

DEFINITION 2.11.4. The set

$$C_x = \{axa^{-1} \mid a \in G\}$$

of elements of that are conjugate to $x \in G$ is called the *conjugacy class* of x.

As a consequence of the fact that conjugacy forms an equivalence relation, any two conjugacy classes are either disjoint or equal.

EXAMPLE 2.11.5. Let $a = r^i$ and $b = r^is$ in D_n. Then we have

$$ar^ja^{-1} = r^i \cdot r^j \cdot r^{-i} = r^j, \qquad a(r^js)a^{-1} = r^i \cdot r^js \cdot r^{-i} = r^{2i+j}s,$$
$$br^jb^{-1} = r^is \cdot r^j \cdot sr^{-i} = r^{-j}, \qquad b(r^js)b^{-1} = r^is \cdot r^js \cdot sr^{-i} = r^{2i-j}s.$$

Therefore, we have $C_{r^i} = \{r^i, r^{-i}\}$ for all $i \in \mathbb{Z}$, while

$$C_s = \{r^{2i}s \mid i \in \mathbb{Z}\},$$

which is all reflections if n is odd, but only half of them if n is even, in which case the remaining conjugacy class is C_{rs}.

DEFINITION 2.11.6. For any $a \in G$, the *conjugation map* is the function defined by

$$\gamma_a \colon G \to G, \qquad \gamma_a(x) = axa^{-1}$$

for $x \in G$.

REMARK 2.11.7. The process of applying a map γ_a to an element of G is referred to as conjugation.

LEMMA 2.11.8. *For $a \in G$, the conjugation map γ_a is an isomorphism.*

PROOF. For $x, y \in G$, we have

$$\gamma_a(x)\gamma_a(y) = (axa^{-1})(aya^{-1}) = a(xy)a^{-1} = \gamma_a(xy).$$

Also, $\gamma_{a^{-1}}$ is the inverse function to γ_a, so γ_a is bijective. □

In particular, we have

$$a(x_1 x_2 \cdots x_s)a^{-1} = ax_1a^{-1} \cdot ax_2a^{-1} \cdot \cdots \cdot ax_sa^{-1} \quad \text{and} \quad (axa^{-1})^{-1} = ax^{-1}a^{-1}$$

for any elements of G.

One very interesting example is conjugation in S_n. We describe this in the case of a cycle.

LEMMA 2.11.9. *Let $\sigma \in S_n$, and let $\tau = (x_1\ x_2\ \cdots\ x_k) \in S_n$ be a k-cycle. Then*

$$\sigma\tau\sigma^{-1} = (\sigma(x_1)\ \sigma(x_2)\ \cdots\ \sigma(x_k)).$$

PROOF. We check this as functions. Let $x \in X_n$. Then

$$\tau(\sigma^{-1}(x)) = \begin{cases} x_{i+1} & \text{if } \sigma^{-1}(x) = x_i, 1 \le i \le k-1 \\ x_1 & \text{if } \sigma^{-1}(x) = x_k \\ \sigma^{-1}(x) & \text{otherwise} \end{cases} = \begin{cases} x_{i+1} & \text{if } x = \sigma(x_i), 1 \le i \le k-1 \\ x_1 & \text{if } x = \sigma(x_k) \\ \sigma^{-1}(x) & \text{otherwise} \end{cases}$$

It follows that

$$\sigma(\tau(\sigma^{-1}(x))) = \begin{cases} \sigma(x_{i+1}) & \text{if } x = \sigma(x_i), 1 \le i \le k-1 \\ \sigma(x_1) & \text{if } x = \sigma(x_k) \\ x & \text{otherwise,} \end{cases}$$

but the latter just the value of the cycle $(\sigma(x_1)\ \sigma(x_2)\ \cdots\ \sigma(x_k))$ on x. □

REMARK 2.11.10. We can use Lemma 2.11.9 to compute the conjugate of any permutation τ by a permutation σ, as τ can be written as a product of cycles, $\tau = \tau_1\tau_2\cdots\tau_s$ and

$$\sigma\tau\sigma^{-1} = \sigma\tau_1\sigma^{-1} \cdot \sigma\tau_2\sigma^{-1} \cdots \sigma\tau_s\sigma^{-1}.$$

EXAMPLE 2.11.11. In S_7, we have

$$(1\ 2\ 3\ 4) \cdot (2\ 3\ 7)(4\ 5) \cdot (1\ 2\ 3\ 4)^{-1} = (3\ 4\ 7)(1\ 5).$$

That is, we have replaced the entries in the permutation $(2\ 3\ 7)(4\ 5)$ that $(1\ 2\ 3\ 4)$ moves to the values it takes them to.

DEFINITION 2.11.12. Write $\tau \in S_n$ as a product of disjoint cycles $\tau = \tau_1\tau_2\cdots\tau_s$ with lengths $2 \le k_1 \le k_2 \le \cdots k_s \le n$. Then $(k_1, k_2, \ldots, k_s)$ is said to be the *cycle type* of τ.

We may use Lemma 2.11.9 to prove the following.

PROPOSITION 2.11.13. *Two elements in S_n are conjugate if and only if they have the same cycle type.*

PROOF. By Lemma 2.11.9 and Remark 2.11.10, any two conjugate permutations must have the same cycle type. On the other hand, we will show that every permutation with cycle type $(k_1,k_2,\dots,k_s)$ is conjugate to a particular permutation α depending only on $(k_1,k_2,\dots,k_s)$. Since conjugacy of elements forms an equivalence relation, we will then have the result.

We first define α. Let $m_i = k_1+k_2+\cdots+k_{i-1}$ for each $1 \le i \le s+1$. In particular, $m_1 = 0$. Define $\alpha_i \in S_n$ by

$$\alpha_i = (m_i+1\ m_i+2\ \cdots\ m_i+k_i).$$

Then the α_i are disjoint cycles, and we set $\alpha = \alpha_1\alpha_2\cdots\alpha_s$.

Now suppose that τ has cycle type $(k_1,k_2,\dots,k_s)$. We must show that τ is conjugate to α. Write

$$\tau = \tau_1\tau_2\cdots\tau_s,$$

where the τ_i are disjoint cycles:

$$\tau_i = (x_{m_i+1}\ x_{m_i+2}\ \cdots\ x_{m_i+k_i}).$$

Now choose any $\sigma \in S_n$ such that $\sigma(i) = x_i$ for each $1 \le i \le m_s$. (For each $m_s < i \le n$, we are free to successively choose the $\sigma(i)$ as i increases to be any values between 1 and n not yet chosen.) Then

$$\sigma\alpha_i\sigma^{-1} = (\sigma(m_i+1)\ \sigma(m_i+2)\ \cdots\ \sigma(m_i+k_i)) = (x_{m_i+1}\ x_{m_i+2}\ \cdots\ x_{m_i+k_i})$$

for each $1 \le i \le s$, so $\sigma\alpha\sigma^{-1} = \tau$, as desired. □

REMARK 2.11.14. In other words, the conjugacy class of a permutation is all permutations with that same cycle type.

2.12. Normal subgroups

We now focus our attention on a very special class of subgroups of a group.

DEFINITION 2.12.1. A subgroup N of a group G is said to be *normal* if $aN = Na$ for every $a \in G$. We also say that N is *normal* in G, and we write $N \trianglelefteq G$ to indicate this.

NOTATION 2.12.2. If N is a proper normal subgroup of a group G, then we write $N \lhd G$.

Of course, the trivial subgroup and the improper subgroup of a group G are, by this definition, normal subgroups. If G is abelian, then every subgroup is normal. We also have the following.

LEMMA 2.12.3. *If H is an index 2 subgroup of a group G, then H is normal in G.*

PROOF. Since G has just two left H-cosets, one of which is H, the other must be the complement of H in G. The same holds for the right H-cosets, hence the result. □

Here are a couple more examples, the second of which we can see from Lemma 2.12.3.

EXAMPLES 2.12.4.

a. We have $\langle r^i \rangle \lhd D_n$ for all $n \ge 3$ and $i \in \mathbb{Z}$. To see this, note that r^i commutes with every r^j, while

$$r^j s \cdot r^i = r^{-i} \cdot r^j s,$$

and $r^{-i} \in \langle r^i \rangle$.

b. We have that $A_n \lhd S_n$ for all $n \geq 2$.

REMARK 2.12.5. If H is a subgroup of G that is not normal, then $aH \neq Ha$ for some $a \in G$. But note that $a \in aH \cap Ha$, and distinct right cosets are disjoint, so aH cannot equal any right coset of G. Therefore, there exists a left coset that is not equal to a right coset.

We can give another characterization of normal subgroups using conjugation. For this, we use the following definition.

DEFINITION 2.12.6. Let H be a subgroup of G. For $a \in G$, the *conjugate subgroup* of H by a is the set of conjugates of a:

$$aHa^{-1} = \{aha^{-1} \mid h \in H\}.$$

That the conjugate subgroup is, in fact, a subgroup is a corollary of Lemma 2.11.8:

COROLLARY 2.12.7. *For $H \leqslant G$ and $a \in G$, we have that $aHa^{-1} \leqslant G$.*

PROOF. We have $aHa^{-1} = \gamma_a(H)$, and the latter is the image of a (sub)group under a homomorphism to G, hence a subgroup of G. □

EXAMPLES 2.12.8.

a. In D_n, we have

$$s\langle r\rangle s^{-1} = \langle r\rangle \quad \text{and} \quad r\langle s\rangle r^{-1} = \langle r^2 s\rangle.$$

b. In S_4, we have

$$(2\,3\,4)\langle(1\,2),(3\,4)\rangle(2\,3\,4)^{-1} = \langle(1\,3),(2\,4)\rangle.$$

LEMMA 2.12.9. *A subgroup N of G is normal if and only if $aNa^{-1} = N$ for all $a \in G$.*

PROOF. Let $a \in G$. The function $\theta_a \colon G \to G$ given by right multiplication by a, i.e., $\theta_a(g) = ga$ is a bijection by the cancellation theorem. Moreover, θ_a restricts to bijections $aNa^{-1} \to aN$ and $N \to Na$, so $aNa^{-1} = N$ if and only if $aN = Na$. □

COROLLARY 2.12.10. *A subgroup N of G is normal if and only if $ana^{-1} \in N$ for all $a \in G$ and $n \in N$.*

PROOF. The only if direction follows from Lemma 2.12.9. On the other hand, the condition $ana^{-1} \in N$ for all $a \in G$ and $n \in N$ clearly implies that $aNa^{-1} \leqslant N$, which we have seen implies $aN \leqslant Na$. But it also means $Na^{-1} \leqslant a^{-1}N$ for all $a \in G$, and this equation for a^{-1} reads $Na \leqslant aN$, which means $aN = Na$, as desired. □

The following proposition gives an extremely useful criterion for a group to be normal.

PROPOSITION 2.12.11. *Let $\phi \colon G \to G'$ be a homomorphism of groups. Then* $\ker\phi$ *is a normal subgroup of G.*

PROOF. Let $a \in G$ and $n \in \ker\phi$. Then we have

$$\phi(ana^{-1}) = \phi(a)\phi(n)\phi(a)^{-1} = \phi(a)\phi(a)^{-1} = e',$$

where e' is the identity of G'. In other words, $ana^{-1} \in N$, so N is normal by Corollary 2.12.10. □

EXAMPLE 2.12.12. The special linear group $\mathrm{SL}_n(\mathbb{R})$ is a normal subgroup of the general linear group $\mathrm{GL}_n(\mathbb{R})$, as it is the kernel of the determinant map.

Here are two other examples.

EXAMPLE 2.12.13. Consider the group

$$\mathrm{Aff}(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \middle| a \in \mathbb{R}^\times, b \in \mathbb{R} \right\},$$

which is a subgroup of $\mathrm{GL}_2(\mathbb{R})$. Then the set

$$N = \left\{ \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \middle| t \in \mathbb{R} \right\}$$

is a normal subgroup of G. To see this is either the following calculation

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} a & at+b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a^{-1} & -a^{-1}b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & at \\ 0 & 1 \end{pmatrix} \in N,$$

or much more simply, that N is the kernel of the restriction of the determinant map to G.

On the other hand, the subgroup

$$H = \left\{ \begin{pmatrix} x & 0 \\ 0 & 1 \end{pmatrix} \mid x \in \mathbb{R}^\times \right\}$$

is not a normal subgroup of G. In fact,

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} ax & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a^{-1} & -a^{-1}b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} x & (1-x)b \\ 0 & 1 \end{pmatrix},$$

and the latter element is not in H if $x \neq 1$ and $b \neq 0$.

EXAMPLE 2.12.14. Let $n \geq 3$. Let $\tau \in S_n$ be a k-cycle with $k \geq 2$. Then $\langle \tau \rangle$ is not normal in S_n unless $n = k = 3$. If $\langle \tau \rangle$ were normal in S_n, then every conjugate of τ would have to be a nontrivial power of τ, of which there are $k-1$. On the other hand, the conjugates of τ are exactly the k-cycles, of which there are

$$\frac{n!}{k(n-k)!} \geq \frac{(n-1)!}{(n-k)!} \geq n-1,$$

which forces $n = k$ in order that $\langle \tau \rangle$ might possibly be normal. But for $n = k$, we have

$$\frac{n!}{k(n-k)!} = (n-1)!,$$

and the latter term is greater than n if $n > 3$, so $n = k = 3$. On the other hand, we have already seen that

$$A_3 = \langle (1\ 2\ 3) \rangle = \langle (1\ 3\ 2) \rangle \lhd S_3.$$

2.13. Quotient groups

The sets of left and right cosets of a normal subgroup are of course the same set, and in this section we prove that this set can be given the structure of a group.

DEFINITION 2.13.1. Let A and B be subsets of a group G. Then we define the *product* of A and B as

$$AB = \{ab \mid a \in A, b \in B\}$$

Moreover, if g is an element of G, we define

$$gA = \{ga \mid a \in A\} \quad \text{and} \quad Ag = \{ag \mid a \in A\}.$$

REMARK 2.13.2. If H is a subgroup of G, then $HH = H$.

THEOREM 2.13.3. *Let N be a normal subgroup of a group G. Then the product of cosets as subsets of G provides a binary operation on G that satisfies*

$$aN \cdot bN = aNbN = abN.$$

Moreover, G/N is a group under this operation.

PROOF. Let $a, b \in G$. Since N is normal in G, we have $Nb = bN$. Therefore, as sets we have

$$aNbN = a(Nb)N = a(bN)N = abNN = abN,$$

as desired. The associativity of the operation is a direct consequence of the associativity of the operation on G, as $(ab)cN = a(bc)N$ for any $a, b, c \in G$. Then $N = eN$ is easily seen to be the identity element of G/N, and the inverse of aN is $a^{-1}N$, since

$$aN \cdot a^{-1}N = aa^{-1}N = N = a^{-1}N \cdot aN.$$

Therefore G/N is a group under this operation. □

DEFINITION 2.13.4. Let N be a normal subgroup of a group G. The *quotient group* of G by N is the group that is the set G/N with the binary operation $aN \cdot bN = abN$ for $a, b \in G$.

EXAMPLES 2.13.5.

a. The set of cosets of $n\mathbb{Z}$ in $\mathbb{Z}$ is a group under addition of cosets, and it is exactly the group $\mathbb{Z}/n\mathbb{Z}$ constructed before, since

$$(a + n\mathbb{Z}) + (b + n\mathbb{Z}) = (a + b) + n\mathbb{Z}$$

by definition.

b. The quotient groups $D_n/\langle r\rangle$ for $n \geq 3$ and S_n/A_n for $n \geq 2$ are all cyclic groups of order 2.

c. Suppose that $n \geq 4$ is even. Let $H = \langle r^2\rangle \lhd D_n$. Then $[D_n : H] = 4$, and the four distinct cosets are H, rH, sH, and rsH. Since the square of each of these cosets is H, we have an isomorphism between D_n/H and the Klein four-group

$$D_n/H \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$$

that takes rH to $(1,0)$ and $sH \to (0,1)$.

REMARK 2.13.6. If N is not normal in G, then G/N is not a group under the product of left cosets. In fact, for $a,b \in N$, the set $aNbN$ will not in general be a left coset. E.g., if one takes $b = a^{-1}$ and a to be such that aNa^{-1} contains an element not in N, then $aNa^{-1}N$ will contain but not equal N, so it is not a left coset.

Moreover, if one simply tries to define $aN \cdot bN = abN$, then the resulting operation is not well-defined, as it depends on the choice of coset representatives. E.g., assuming it were well-defined and again taking a and b as above, we can find $n \in N$ such that $ana^{-1} \notin N$, so

$$N = aN \cdot a^{-1}N = anN \cdot a^{-1}N = ana^{-1}N \neq N,$$

which contradicts well-definedness.

Note that the function $G \to G/N$ that takes $g \in G$ to its N-coset gN is a homomorphism by definition of the quotient group. We give it a name.

DEFINITION 2.13.7. Let N be a normal subgroup of a group G. The *quotient map* $\pi_N \colon G \to G/N$ is the homomorphism defined by $\pi_N(g) = gN$ for $g \in G$.

REMARK 2.13.8. The kernel of the quotient map $\pi_N \colon G \to G/N$ is N.

COROLLARY 2.13.9. *A subgroup N of G is normal if and only if there exists a group G' and a group homomorphism $\phi \colon G \to G'$ such that $N = \ker\phi$.*

We end with the following result on the subgroups of quotient groups.

PROPOSITION 2.13.10. *Let G be a group and N be a normal subgroup of G. Then the subgroups of G/N are exactly the quotient groups H/N, where H is a subgroup of G containing N. Moreover, such a subgroup H of G is normal in G if and only if H/N is normal in G/N.*

PROOF. We first note that if H is a subgroup of G containing N, then N is normal in H, so we may form the quotient group H/N. Its binary operation agrees with the restriction of the operation on G/N (multiplication of N-cosets), so it is a subgroup of G/N.

Conversely, if Q is a subgroup of G/N, then set

$$H = \{h \in G \mid hN \in Q\}.$$

Then $nN = N \in Q$, in that it is the identity element of G/N and Q is a subgroup, so $N \subseteq H$. That H is a subgroup of G follows directly from the fact that Q is a subgroup of G/N, since if $hN, kN \in Q$, then $hN \cdot (kN)^{-1} \in Q$, so $hk^{-1}N \in Q$, which means that $hk^{-1} \in H$.

Finally, let H be a subgroup of G containing N. Then, for $a \in G$, we have

$$aN \cdot (H/N) = \{ahN \mid h \in H\} = aH$$

and

$$(H/N) \cdot aN = \{Nha \mid h \in H\} = Ha,$$

so $H/N \trianglelefteq G/N$ if and only if $H \trianglelefteq G$. □

THEOREM 2.13.11 (First Isomorphism Theorem). *Let $\phi \colon G \to G'$ be a homomorphism of groups. Then the function*

$$\bar{\phi} \colon G/\ker\phi \xrightarrow{\sim} \operatorname{im}\phi, \qquad \bar{\phi}(a\ker\phi) = \phi(a)$$

for $a \in G$ is a well-defined group isomorphism.

PROOF. First, note that if $a,b \in G$ are such that $a\ker\phi = b\ker\phi$, then $a = bk$ for some $k \in \ker\phi$, so

$$\bar{\phi}(a\ker\phi) = \phi(bk) = \phi(b)\phi(k) = \phi(b) = \bar{\phi}(b\ker\phi),$$

and hence $\bar{\phi}$ is well-defined. Moreover, if $\bar{\phi}(a\ker\phi) = 0$, then $\phi(a) = 0$, so $a \in \ker\phi$, and therefore we have that $\bar{\phi}$ is injective. Since ϕ has image $\operatorname{im}\phi$, so does $\bar{\phi}$, and hence $\bar{\phi}$ is surjective by definition. □

REMARK 2.13.12. We have that $\phi = \iota_{\operatorname{im}\phi} \circ \bar{\phi} \circ \pi_{\ker\phi}$, as is represented in the following diagram:

$$\begin{tikzcd}
G \arrow[rrr, "\phi"] \arrow[dr, "\pi_{\ker\phi}"'] & & & G'. \\
& G/\ker\phi \arrow[r, "\bar{\phi}"] & \operatorname{im}\phi \arrow[ur, "\iota_{\operatorname{im}\phi}"'] &
\end{tikzcd}$$

EXAMPLE 2.13.13. The determinant map $\det\colon \mathrm{GL}_n(\mathbb{R}) \to \mathbb{R}^\times$ induces an isomorphism

$$\overline{\det}\colon \mathrm{GL}_n(\mathbb{R})/\mathrm{SL}_n(\mathbb{R}) \xrightarrow{\sim} \mathbb{R}^\times.$$

To give another example, we make the following definition.

DEFINITION 2.13.14. Let $G_1, G_2, \ldots, G_k$ be groups. The ith *projection map* is the surjective homomorphism

$$\pi_i\colon \prod_{j=1}^k G_j \to G_i, \qquad \pi_i(g_1, g_2, \ldots, g_k) = g_i.$$

REMARK 2.13.15. The projection map π_i of Definition 2.13.14 has kernel

$$\ker\pi_i = \left\{ (g_1, g_2, \ldots, g_k) \in \prod_{j=1}^k G_j \mid g_j = e_j \text{ for all } j \neq i \right\},$$

where e_i is the identity element of G_i. By the first isomorphism theorem, π_i induces an isomorphism

$$\bar{\pi}_i\colon \Big(\prod_{i=1}^k G_j\Big)/\ker\pi_i \xrightarrow{\sim} G_i.$$

For instance, if $k = 2$ and $i = 1$, we can think of $\bar{\pi}_i$ as an isomorphism

$$\frac{G_1 \times G_2}{\{e_1\} \times G_2} \xrightarrow{\sim} G_1.$$

EXAMPLE 2.13.16. We have

$$\frac{\mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}}{\langle 2 \rangle \times \langle 2 \rangle} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}.$$

To see this, define a map $\phi\colon \mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ by $\phi(a,b) = (a,b)$. Then ϕ is surjective with kernel $\langle 2 \rangle \times \langle 2 \rangle$, so the first isomorphism theorem applied to ϕ provides the isomorphism.

CHAPTER 3

Ring theory

3.1. Rings

In this section, we define rings and fields. These are sets with two binary operations, known as addition and multiplication.

DEFINITION 3.1.1. Let R be a set with a pair $(+,\cdot)$ of binary operations. We say that R satisfies the *left distributive law* (with respect to $+$ and $\cdot$) if

$$a\cdot(b+c) = (a\cdot b)+(a\cdot c)$$

for all $a,b,c \in R$, and we say that R satisfies the *right distributive law* if

$$(a+b)\cdot c = (a\cdot c)+(b\cdot c)$$

for all $a,b,c \in R$.

The distributive law being one of the standard axioms of arithmetic, it is satisfied by many common objects, such as $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, and so on. We give one less standard example.

EXAMPLE 3.1.2. The set $\mathrm{Maps}(\mathbb{R},\mathbb{R})$ satisfies the left and right distributive laws with respect to the pair of operations $(+,\cdot)$. It satisfies the right distributive law with respect to $(+,\circ)$ and $(\cdot,\circ)$, where $\circ$ is composition.

We now define a ring.

DEFINITION 3.1.3. A set R with a pair $(+,\cdot)$ of binary operations is a *ring* if

i. R is an abelian group under $+$,

ii. the binary operation $\cdot$ is associative,

iii. R has an identity element 1 under $\cdot$, and

iv. R satisfies the left and right distributive laws.

REMARK 3.1.4. When $+$ and $\cdot$ are used to denote the binary operations of a ring, we refer $+$ as addition and $\cdot$ as multiplication. Unless otherwise stated, the operations of R will be denoted $+$ and $\cdot$.

REMARK 3.1.5. As in the case of groups, we often write ab for $a\cdot b$ for a,b in a ring R. We also use $a\cdot b+c$ to denote $(a\cdot b)+c$ for $a,b,c \in R$.

EXAMPLES 3.1.6.

a. The sets $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all rings with respect to the usual operations of addition and multiplication.

b. The sets $\text{Maps}(\mathbb{R},\mathbb{R})$ and $M_n(\mathbb{R})$ for $n \geq 1$ are also rings with respect to addition and multiplication.

c. The set $n\mathbb{Z}$ for $n \geq 1$ is a ring with respect to addition and multiplication.

d. The set $\mathbb{Z}/n\mathbb{Z}$ is a ring with respect to its operations of addition and multiplication.

REMARK 3.1.7. Since the first binary operation on a ring R is denoted $+$, the identity element is denoted 0 as usual, and the additive inverse of $a \in R$ is denoted $-a$. The sum of n copies of a is denoted na for $n \geq 1$, and $-(na)$ is also denoted $-na$.

We have the following properties in any ring.

LEMMA 3.1.8. *Let R be a ring, and let $a, b \in R$. Then we have*

a. $0 \cdot a = a \cdot 0 = 0$,

b. $a \cdot (-b) = (-a) \cdot b = -ab$, *and*

c. $(-a) \cdot (-b) = ab$.

PROOF.

a. We have

$$0 \cdot a + b \cdot a = (0 + b) \cdot a = b \cdot a$$

by the right distributive law and the fact that 0 is an additive identity. Therefore, the Cancellation theorem tells us that $0 \cdot a = 0$. Similarly, $a \cdot 0 = 0$ using the left distributive law instead of the right.

b. We have

$$a \cdot (-b) + ab = a \cdot (-b + b) = a \cdot 0 = 0$$

by the left distributive law, the definition of the additive inverse, and part a. The other equality is similar.

c. This follows from part b, which tells us that

$$(-a) \cdot (-b) = -(a \cdot (-b)) = -(-(ab)) = ab.$$

□

On a set with one element, there is only one possible binary operation, and using it as both addition and multiplication turns that set into a ring.

DEFINITION 3.1.9. The *zero ring* is the ring $\{0\}$. We say that a ring R is a nonzero ring if R has more than one element.

That a ring R has an identity under $\cdot$ is to say exactly that there is an element $1 \in R$ with $1 \cdot a = a \cdot 1 = a$ for all $a \in R$. By Lemma 2.1.4, the multiplicative identity in a ring is unique.

EXAMPLES 3.1.10. The rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\text{Maps}(\mathbb{R},\mathbb{R})$, $M_n(\mathbb{R})$ for $n \geq 1$, $\mathbb{Z}/n\mathbb{Z}$ for $n \geq 1$, and $\{0\}$ are all rings with unity. However, $n\mathbb{Z}$ is not a ring for $n \geq 2$.

REMARK 3.1.11. One easily checks that $(na) \cdot (mb) = (nm)ab$ for $n, m \in \mathbb{Z}$ and $a, b \in R$ for any ring R. One has $(n \cdot 1) \cdot (m \cdot 1) = nm \cdot 1$. We often denote $n \cdot 1$ by n, though we remark that it is possible that $n \cdot 1 = m \cdot 1$ for $n \neq m$, as will happen in any finite ring, for instance.

REMARK 3.1.12. If R is a ring with $1 = 0$, then $x = 1 \cdot x = 0 \cdot x = 0$ for all $x \in R$, so R is the zero ring.

We now introduce the notion of a subring of a ring, which does not play quite as prominent of a role in ring theory as does the notion of a subgroup of a group in group theory.

DEFINITION 3.1.13. A *subring* S of a ring R is a subset of R that is a ring with respect to the restrictions to S of the binary operations of addition and multiplication on R.

We leave it to the reader to check the following.

LEMMA 3.1.14. *A subset S of a ring R is a subring if it is closed under the operations of addition and multiplication on R, contains* 0 *and* 1*, and contains $-a$ for all $a \in S$.*

Clearly, the property of being a subring is a transitive one.

EXAMPLES 3.1.15.

a. The set $\{0\}$ is a subring of any ring.

b. The ring $n\mathbb{Z}$ is not a subring of $\mathbb{Z}$, as it does not contain 1.

c. The ring $\mathbb{Z}$ is a subring of $\mathbb{Q}$, which is in turn a subring of $\mathbb{R}$, which is in turn a subring of $\mathbb{C}$.

Most of the study of ring theory is focused on commutative rings.

DEFINITION 3.1.16. A ring R is a *commutative ring* if multiplication on R is commutative. We then say that the ring R is *commutative*.

DEFINITION 3.1.17. A ring R that is not commutative is a *noncommutative ring*.

EXAMPLES 3.1.18. The rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathrm{Maps}(\mathbb{R},\mathbb{R})$, $\mathbb{Z}/n\mathbb{Z}$ for $n \geq 1$, and $\{0\}$ are all commutative rings. However, $M_n(\mathbb{R})$ is a noncommutative ring for all $n \geq 2$.

The notion of a field is really just a special case of the notion of a ring, but it is an important one.

DEFINITION 3.1.19. A *field* is a nonzero commutative ring for which every nonzero element has a multiplicative inverse.

In other words, a field is a nonzero commutative ring for which the nonzero elements form a group under multiplication (in fact, an abelian group).

DEFINITION 3.1.20. A *subfield* of a field F is a subring of F that is a field.

EXAMPLES 3.1.21.

a. The rings $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields. Of course, $\mathbb{Q}$ is a subfield of $\mathbb{R}$ and $\mathbb{C}$, and $\mathbb{R}$ is a subfield of $\mathbb{C}$.

b. The ring $\mathbb{Z}$ is not a field.

The analogous object to a field in the more general theory of possibly noncommutative rings is known as a division ring.

DEFINITION 3.1.22. A *division ring* (or *skew field*) D is a nonzero ring such that every nonzero element is invertible under multiplication.

Clearly, all fields are division rings. As with fields, we have multiplicative groups of division rings, which no longer need be abelian.

DEFINITION 3.1.23. The group of nonzero elements in a division ring D is known as the *multiplicative group* of D and is denoted $D^\times$.

We end this section with one example of a noncommutative division ring.

DEFINITION 3.1.24. The ring of quaternions $\mathbb{H}$ is the set of distinct elements $a+bi+cj+dk$ with $a,b,c,d \in \mathbb{R}$, together with addition defined by

$$(a+bi+cj+dk)+(a'+b'i+c'j+d'k) = (a+a')i+(b+b')j+(c+c')k$$

and multiplication defined by

$$\begin{aligned}(a+bi+cj+dk)\cdot(a'+b'i+c'j+d'k) &= (aa'-bb'-cc'-dd') \\ &+ (ab'+ba'+cd'-dc')i+(ac'-bd'+ca'+db')j+(ad'+bc'-cb'+da')k\end{aligned}$$

for $a,b,c,d,a',b',c',d' \in \mathbb{R}$.

REMARK 3.1.25. The ring $\mathbb{H}$ is an $\mathbb{R}$-vector space with basis 1, i, j, k, where

$$\alpha\cdot(a+bi+cj+dk) = \alpha a+(\alpha b)i+(\alpha c)j+(\alpha d)k$$

for $\alpha,a,b,c,d \in \mathbb{R}$. Note that we have $ij = k = -ji$, $jk = i = -kj$, $ki = j = -ik$, and $i^2 = j^2 = k^2 = -1$ in $\mathbb{H}$.

THEOREM 3.1.26. *The quaternion algebra is a division ring.*

PROOF. We give only a sketch. Distributivity is a direct consequence of the definitions of the operations of addition and multiplication. In fact, it is also easy to see that $\alpha\cdot xy = x\cdot(\alpha y) = (\alpha x)\cdot y$ for $\alpha \in \mathbb{R}$ and $x,y \in \mathbb{H}$. Using the distributive law and the latter fact, associativity of multiplication follows from a check of associativity on the subset $\{i,j,k\}$ of $\mathbb{H}^\times$. Finally, any nonzero $a+bi+cj+dk \in \mathbb{H}$ has inverse

$$(a+bi+cj+dk)^{-1} = (a^2+b^2+c^2+d^2)^{-1}(a-bi-cj-dk),$$

so $\mathbb{H}$ is a division ring. □

3.2. Families of rings

In this section, we consider various sorts, or families, of rings one can construct out of other rings. We begin with matrix rings.

DEFINITION 3.2.1. If R is a nonzero ring, the matrix ring $M_n(R)$ consisting of n-by-n matrices with entries in R is the set with the addition $(a_{ij})+(b_{ij}) = (a_{ij}+b_{ij})$ and multiplication

$$(a_{ij})\cdot(b_{ij}) = \left(\sum_{k=1}^{n} a_{ik}b_{kj}\right).$$

We leave the proof of the following to the reader.

LEMMA 3.2.2. *Let R be a ring and $n \geq 1$. Then $M_n(R)$ is a ring.*

LEMMA 3.2.3. *The ring $M_n(R)$ is noncommutative if R is a ring and $n \geq 2$.*

PROOF. Suppose $n = 2$. Let $A = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$. Then $AB = I$, the identity matrix in $M_2(\mathbb{R})$, while $BA = 0$. The general case follows from the case $n = 2$ by taking matrices that contain the same entries as A and B in their upper lefthand corners and are zero in all other entries. □

Another important class of rings is the polynomial rings.

DEFINITION 3.2.4. Let R be a ring, and fix an indeterminate (i.e., a symbol) x. The polynomial ring $R[x]$ with R-coefficients is the set of finite formal (i.e., two are different if they are written differently) sums of powers of x with coefficients in R, i.e.,

$$R[x] = \left\{ \sum_{i=0}^{\infty} a_i x^i \mid a_i \in R \text{ for all } i \geq 0, a_i = 0 \text{ for all } i > N \text{ for some } N \geq 0 \right\}.$$

together with the binary operations of addition and multiplication given by

$$\sum_{i=0}^{\infty} a_i x^i + \sum_{j=0}^{\infty} b_j x^j = \sum_{i=0}^{\infty} (a_i + b_i) x^i,$$

$$\left(\sum_{i=0}^{\infty} a_i x^i \right) \cdot \left(\sum_{j=0}^{\infty} b_j x^j \right) = \sum_{k=0}^{\infty} \left(\sum_{i=0}^{k} a_i b_{k-i} \right) x^k.$$

An element $f = \sum_{i=0}^{\infty} a_i x^i$ of $R[x]$ is called a polynomial, the a_i are referred to as coefficients, and x is called a variable.

REMARK 3.2.5. If $a_i = 0$ for all $i > N$, then we more commonly write $\sum_{i=0}^{N} a_i x^i$ for $f = \sum_{i=0}^{\infty} a_i x^i$. We will also sometimes write

$$f = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$$

identifying x^0 with "1" and x^1 with "x".

DEFINITION 3.2.6. The degree $\deg f$ of a nonzero polynomial $f = \sum_{i=0}^{\infty} a_i x^i \in R[x]$ is the smallest integer N such that $a_i = 0$ for all $i > N$. When needed, we consider the degree of 0 to be $-\infty$.

REMARK 3.2.7. A polynomial of degree 0 is said to be constant, a polynomial of degree 1 is linear, a polynomial of degree 2 is quadratic, followed by cubic, quartic, quintic, and so forth.

DEFINITION 3.2.8. If f is a polynomial of degree $n \geq 0$, then its leading coefficent is the coefficient of x^n in f. If constant coefficient is the coefficient of $x^0 = 1$.

EXAMPLE 3.2.9. The polynomials $2 + 3x - x^2$ and $1 + x$ are elements of $\mathbb{Z}[x]$. One has, as usual,

$$(2 + 3x - x^2) \cdot (1 + x) = 2 + (2 + 3)x + (3 - 1)x^2 - x^3 = 2 + 5x + 2x^2 - x^3.$$

The following is a direct consequence of the definitions of addition and multiplication in polynomial rings.

LEMMA 3.2.10. *Let R be a ring, and let $f,g \in R[x]$ be polynomials. Then* $\deg fg \leq \deg f \cdot \deg g$. *Moreover, we have*

$$\deg(f+g) \leq \max\{\deg f, \deg g\},$$

and equality holds in the last statement if $\deg f \neq \deg g$.

DEFINITION 3.2.11. The polynomials $a = a + 0 \cdot x + 0 \cdot x^2 + \cdots$ for $a \in R$ are referred to as constant polynomials. These are exactly 0 and the polynomials of degree 0. The set of constant polynomials forms a subring of $R[x]$, which we also denote R.

We leave it to the reader to check the following.

LEMMA 3.2.12. *Let R be a ring. Then the polynomial ring $R[x]$ is in fact a ring.*

REMARK 3.2.13. The ring $R[x]$ is commutative if and only if R is commutative. The 1 is a multiplicative identity in R, then 1 is a multiplicative identity in $R[x]$ as well.

We may also consider polynomial rings in several variables.

DEFINITION 3.2.14. Let $n \geq 1$ and $x_1, x_2, \ldots, x_n$ be indeterminates. The polynomial ring in n variables over a ring R is defined to be

$$R[x_1, x_2, \ldots, x_n] = (((R[x_1])[x_2]) \cdots)[x_n].$$

We write an element of this ring as

$$\sum_{i_1=0}^{N_1} \sum_{i_2=0}^{N_2} \cdots \sum_{i_n=0}^{N_n} a_{i_1 i_2 \ldots i_N} x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n},$$

where the coefficients lie in R. The elements $x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$ are called monomials.

We will see below that this construction is independent, up to isomorphism, of the ordering of the variables.

REMARK 3.2.15. In multiplying in $R[x_1, x_2, \ldots, x_n]$, the variables x_i all commute with each other and the elements of R. A quantity such as $x_2 x_1 x_2$ equals $x_1 x_2^2$.

EXAMPLE 3.2.16. In the ring $\mathbb{Z}[x,y]$, we have polynomials like $x^2 + 2xy$ and $1 - x + y$, and we have

$$(x^2 + 2xy)(1 - x + y) = x^2 + 2xy - x^3 + x^2y + 2xy^2.$$

Finally, we consider direct products.

DEFINITION 3.2.17. Let I be an indexing set, and let $\{R_i \mid i \in I\}$ be a nonempty collection of rings. Then the direct product $\prod_{i \in I} R_i$ of the R_i over $i \in I$ is the binary structure is the direct product of the sets R_i together with the binary operations of coordinate-wise addition and multiplication. If $I = X_n = \{1, 2, \ldots, n\}$, we write

$$\prod_{i \in I} R_i = R_1 \times R_2 \times \cdots \times R_n.$$

That the direct product of rings is a ring is a simple consequence of its definition, and we state it without proof.

LEMMA 3.2.18. *Any direct product of rings is a ring.*

REMARKS 3.2.19. Let $\{R_i \mid i \in I\}$ be a nonempty collection of rings, and set $R = \prod_{i \in I} R_i$.

a. The ring R is commutative if and only if each R_i is commutative.

b. The zero element of R is the element $(0)_{i \in I}$.

c. The element $(1)_{i \in I}$ is the multiplicative identity in R.

d. The element e_i which is 0 in every coordinate but the ith, where it is 1, satisfies $e_i^2 = e_i$, but e_i is not the multiplicative identity of R (unless I has only one element).

EXAMPLE 3.2.20. If R is any ring, then R^n is the product of n copies of R.

3.3. Units

Not all rings with unity are fields, but one can still ask which elements are invertible under multiplication. These elements are known as units.

DEFINITION 3.3.1. A unit in a ring is a nonzero element $u \in R$ such that u has a multiplicative inverse in R. We also say that u is invertible.

EXAMPLES 3.3.2.

a. The element 1 is a unit in every nonzero ring.

b. The units in a field F are the elements of $F^\times$.

c. The only units in $\mathbb{Z}$ are 1 and -1.

PROPOSITION 3.3.3. *The units in a nonzero ring R with unity form a group under multiplication.*

PROOF. Let $R^\times$ denote the set of units in R. If $u, v \in R^\times$, then let $u', v' \in R^\times$ be multiplicative inverses to u and v respectively. We have

$$uv \cdot (v'u') = 1 = (v'u') \cdot uv,$$

so multiplication is a binary operation on $R^\times$, which we already know to be associative. Clearly, 1 is a unit and an identity in $R^\times$, and by definition, every unit has an inverse in $R^\times$, so $R^\times$ is a group. □

DEFINITION 3.3.4. The group of units in a nonzero ring R with unity is denoted $R^\times$.

REMARK 3.3.5. If F is a field, then its unit group and its multiplicative group coincide, and hence the notation $F^\times$ for both is unambiguous.

EXAMPLE 3.3.6. The group of units in $M_n(R)$ for a ring R with unity is its subset $\mathrm{GL}_n(R)$ of invertible matrices. E.g., if $R = \mathbb{R}$, then these are the matrices with nonzero determinant.

EXAMPLE 3.3.7. If $R = \prod_{i\in I} R_i$ is a direct product of rings R_i with unity over an indexing set I, then

$$R^\times = \prod_{i\in I} R_i^\times.$$

PROPOSITION 3.3.8. *The units in $\mathbb{Z}/n\mathbb{Z}$ for $n \geq 1$ are exactly the images of those $i \in \mathbb{Z}$ relatively prime to n.*

PROOF. Let $i \in \mathbb{Z}$. By Proposition 2.3.14, we have $\langle i\rangle = \langle \gcd(i,n)\rangle$ as subgroups of $\mathbb{Z}/n\mathbb{Z}$. The set of ij with $j \in \mathbb{Z}/n\mathbb{Z}$ are exactly the elements of $\langle i\rangle$. Therefore, i is a unit in $\mathbb{Z}/n\mathbb{Z}$ if and only if 1 is an integer multiple of $\gcd(i,n)$ in $\mathbb{Z}/n\mathbb{Z}$. Since $\gcd(i,n)$ is a divisor of n, this can and will only happen if $\gcd(i,n) = 1$, which is to say that i is relatively prime to n. □

COROLLARY 3.3.9. *The group $(\mathbb{Z}/n\mathbb{Z})^\times$ has order $\phi(n)$, where ϕ is the Euler ϕ-function.*

COROLLARY 3.3.10. *For $n \geq 1$, the ring $\mathbb{Z}/n\mathbb{Z}$ is a field if and only if n is prime.*

We now have the following corollaries by the corollary of Lagrange's theorem that the order of an element of a group divides the order of the group. What is remarkable is that they are nonobvious statements of simple arithmetic.

COROLLARY 3.3.11 (Euler's theorem). *Let $n \geq 1$. Then*

$$a^{\phi(n)} \equiv 1 \bmod n$$

for every $a \in \mathbb{Z}$ relatively prime to n.

Note that every nonzero element of $\mathbb{Z}/p\mathbb{Z}$ is relatively prime to p. Hence we also also have the following special case of Euler's theorem.

COROLLARY 3.3.12 (Fermat's little theorem). *Let p be a prime number. Then*

$$a^{p-1} \equiv 1 \bmod p$$

for every $a \in \mathbb{Z}$ not divisible by p.

These raise the following questions. What is the order of a unit in $\mathbb{Z}/n\mathbb{Z}$? We know it to be a divisor of $\phi(n)$, but is there a simple formula for it in terms of a and n? This is one of many questions in the field of mathematics known as number theory. Let us give a few examples of arithmetic in $\mathbb{Z}/n\mathbb{Z}$.

EXAMPLE 3.3.13. Suppose we wish to calculate 3^{362} in $\mathbb{Z}/11\mathbb{Z}$. Fermat's little theorem tells us that $3^{10} \equiv 1 \bmod 11$, so

$$3^{362} \equiv (3^{10})^{36} 3^2 \equiv 3^2 \equiv 9 \bmod 11.$$

In other words, $3^{10} = 9$ in $\mathbb{Z}/11\mathbb{Z}$.

EXAMPLE 3.3.14. What is the order of 2 in $(\mathbb{Z}/101\mathbb{Z})^\times$? Since 101 is prime, the order of 2 must be a divisor of 100. We have $2^5 < 101$, and $2^{10} = 1024 \equiv 14 \bmod 101$. Moreover, we have

$$2^{20} \equiv (14)^2 \equiv 196 \equiv -6 \bmod 101,$$
$$2^{25} = 2^{20}2^5 \equiv -6 \cdot 32 \equiv -192 \equiv 10 \bmod 101,$$
$$2^{50} \equiv 10^2 \equiv -1 \bmod 101.$$

Therefore, the order of 2 in $\mathbb{Z}/101\mathbb{Z}$ must be 100.

3.4. Integral domains

DEFINITION 3.4.1. A *left* (resp., *right*) *zero divisor* in a ring R is a nonzero element $a \in R$ such that there exists a nonzero element $b \in R$ with $ab = 0$ (resp., $ba = 0$). A *zero divisor* in a ring R is an element that is either a left or a right zero divisor.

REMARK 3.4.2. Note that 0 is never considered to be a zero divisor (at least under our conventions). In fact, 1 is never a zero divisor either, as $1 \cdot b = b$ for all $b \in R$.

EXAMPLE 3.4.3. The ring $M_2(\mathbb{R})$ has zero divisors. For instance, we have

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = 0.$$

EXAMPLE 3.4.4. If $R = R_1 \times R_2$ for some nonzero rings R_1 and R_2, then R has zero divisors, since if $a \in R_1$ is nonzero and $b \in R_2$ is nonzero, we have $(a,0) \cdot (0,b) = (0,b) \cdot (a,0) = 0$. For instance, $\mathbb{Z}^n$ has zero divisors for $n \geq 2$, though $\mathbb{Z}$ does not.

One might ask for a ring that contains a left zero divisor that is not a right zero divisor. For this, let us make the following general definition.

DEFINITION 3.4.5. Let A be an abelian group under addition. The *endomorphism ring* of A is the set

$$\mathrm{End}(A) = \{f \colon A \to A \mid f \text{ is a group homomorphism}\}$$

under addition and composition of functions.

REMARK 3.4.6. If A is an abelian group, then $\mathrm{End}(A)$ is a ring, with 1 being the identity function on A. In general, $\mathrm{End}(A)$ may be a noncommutative ring.

EXAMPLE 3.4.7. Let $A = \prod_{i=1}^{\infty} \mathbb{Z}$, an abelian group under addition. Define $L, R \in \mathrm{End}(A)$ by

$$L(a_1,a_2,a_3,\ldots) = (a_2,a_3,a_4,\ldots) \quad \text{and} \quad R(a_1,a_2,a_3,\ldots) = (0,a_1,a_2,\ldots).$$

Moreover, let $M \in \mathrm{End}(A)$ be defined by

$$M(a_1,a_2,a_3,\ldots) = (a_1,0,0,\ldots).$$

Then

$$LM(a_1,a_2,a_3,\ldots) = L(a_1,0,0,\ldots) = 0 \quad \text{and} \quad MR(a_1,a_2,a_3,\ldots) = M(0,a_1,a_2,\ldots) = 0,$$

so L is a left zero divisor and R is a right zero divisor. On the other hand,

$$LR(a_1,a_2,a_3,\ldots) = L(0,a_1,a_2,\ldots) = (a_1,a_2,a_3,\ldots),$$

so $LR = 1$. Therefore, L cannot be a left zero divisor, for if $XL = 0$ for some $X \in \mathrm{End}(A)$, then $0 = (XL)R = X(LR) = X$. Similarly, R is not a right zero divisor.

EXAMPLE 3.4.8. In the ring $\mathbb{Z}/6\mathbb{Z}$, the elements 2, 3, and 4 are zero divisors, since $2 \cdot 3 = 3 \cdot 4 = 0$.

More generally, we have the following.

LEMMA 3.4.9. *For $n \geq 1$, the zero divisors in $\mathbb{Z}/n\mathbb{Z}$ are exactly its nonzero elements that are not relatively prime to n.*

PROOF. Let $\bar{a} \in \mathbb{Z}/n\mathbb{Z}$ be nonzero, and let $b = n/\gcd(a,n)$. Then $\bar{a}\bar{b} = 0$, and we know that $\bar{b} \neq 0$ if and only if $\gcd(a,n) \neq 1$. On the other hand, if $\bar{a}\bar{b} = 0$, then ab is a multiple of n, so b is a multiple of $n/\gcd(a,n)$. Therefore, a is a zero divisor if and only if $\gcd(a,n) \neq 1$, which occurs if and only if a is not relatively prime to n □

As a corollary, if p is a prime number, then $\mathbb{Z}/p\mathbb{Z}$ has no zero divisors. In fact, we shall see momentarily that every field has no zero divisors.

DEFINITION 3.4.10. A nonzero commutative ring R with unity is called an *integral domain* if R contains no zero divisors.

LEMMA 3.4.11. *Every field is an integral domain.*

PROOF. Let F be a field, and let $a \in F$ be such that there exists a nonzero element $b \in F$ with $ab = 0$. Then $0 = (ab)b^{-1} = a$. □

By definition, any subring of an integral domain is also an integral domain.

EXAMPLES 3.4.12. The fields $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{Z}/p\mathbb{Z}$ for any prime p are all integral domains. That $\mathbb{Z}$ is an integral domain is either an easy check or the fact that it is a subring of $\mathbb{Q}$ with unity. Since $\mathbb{Z}/n\mathbb{Z}$ contains zero divisors for composite $n \geq 1$, it is not an integral domain.

PROPOSITION 3.4.13. *Let R be an integral domain. Then $R[x]$ is an integral domain. Moreover, if $f,g \in R[x]$ are nonzero, then $\deg fg = \deg f + \deg g$, and the units in $R[x]$ are exactly the units in R.*

PROOF. Let $f,g \in R[x]$ be nonzero polynomials of degree N and M respectively. Write $f = \sum_{i=0}^{N} a_i x^i$ and $g = \sum_{j=0}^{M} b_j x^j$. Then

$$fg = \sum_{k=0}^{N+M} c_k x^k, \qquad c_k = \sum_{i=0}^{k} a_i b_{k-i}.$$

If $0 \leq i \leq N+M$, then $a_i = 0$ if $i > N$ and $b_{M+N-i} = 0$ if $i < N$, so $c_{N+M} = a_N b_M$. Since R is an integral domain, we then have $c_{N+M} \neq 0$, so $fg \neq 0$. Therefore, we have $\deg fg = N+M$. If $fg = 1$, then this forces $N = M = 0$, and therefore $f = a_0$, $g = b_0$, and $a_0 b_0 = 1$, which means that $f \in R^\times$. □

One particularly nice use of integral domains is that they obey cancellation laws.

LEMMA 3.4.14. *Let R be an integral domain, and let $a,b,c \in R$ be such that $ab = ac$. Then either $a = 0$ or $b = c$.*

PROOF. If $ab = ac$, then $a(b-c) = 0$ by the distributive law (and Lemma 3.1.8), so as R contains no zero divisors, at least one of a and $b-c$ must be 0. □

We have already seen that $\mathbb{Z}/n\mathbb{Z}$ is an integral domain if and only if n is prime, and so if and only if $\mathbb{Z}/n\mathbb{Z}$ is a field. We have the following stronger result.

THEOREM 3.4.15. *If R is a finite integral domain, then R is a field.*

PROOF. Let $a \in R$ be nonzero. Lemma 3.4.14 tells us that the elements ab with $b \in R$ are all distinct. Since there are then $|R|$ of them, the set $\{ab \mid b \in R\}$ is R itself. In particular, there exists $b \in R$ with $ab = 1$, proving that a has a multiplicative inverse. □

Finally, we introduce the notion the characteristic of a ring.

DEFINITION 3.4.16. Let R be a ring. The *characteristic* $\mathrm{char}(R)$ of R is the smallest $n \geq 1$ such that $na = 0$ for all $a \in R$ if such an n exists, and otherwise we set $\mathrm{char}(R) = 0$.

EXAMPLES 3.4.17.

a. The ring $\mathbb{Z}/n\mathbb{Z}$ has characteristic n, while $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ all have characteristic 0.

b. The characteristic of $M_n(R)$ for a ring R is equal to the characteristic of R for every $n \geq 1$.

LEMMA 3.4.18. *The characteristic of a nonzero ring R with unity is the smallest $n > 1$ such that $n = 0$ in R if such an n exists, and is 0 otherwise.*

PROOF. We cannot have $1 \cdot a = 0$ unless $a = 0$, so $\mathrm{char}(R) \neq 1$ as R is nonzero. Recall that $n \in R$ is considered to be $n \cdot 1$. If $n = 0$ in R, then clearly $na = 0$ for all $a \in R$. On the other hand, that $n = 0$ is the special case of $na = 0$ with $a = 1$. If $n = n \cdot 1 \neq 0$ for all $n \geq 1$, then by definition, we have $\mathrm{char}(R) = 0$. □

PROPOSITION 3.4.19. *The characteristic of an integral domain is either 0 or prime.*

PROOF. We employ Lemma 3.4.18. If R is an integral domain and $n = 0$ in R for some composite $n > 1$, then $n = mm' = 0$ for some prime m and m' dividing n, which by the nonexistence of zero divisors implies that either m or m' is zero. In other words, the smallest $n > 1$ with $n \neq 0$ in R cannot be composite, so must be prime. □

3.5. Ring homomorphisms

In this section, we introduce the notion of a ring homomorphism, which is a function from one ring to another that is compatible with both addition and multiplication: in other words, it is a homomorphism of binary structures both for $(R,+)$ and for $(R,\cdot)$.

DEFINITION 3.5.1. Let R and S be rings. A function $\phi \colon R \to S$ is a ring homomorphism if $\phi(1) = 1$ and it satisfies

$$\phi(a+b) = \phi(a) + \phi(b) \quad \text{and} \quad \phi(ab) = \phi(a)\phi(b)$$

for all $a, b \in R$.

We give some examples of ring homomorphisms.

EXAMPLES 3.5.2.

a. The reduction map $\phi_n \colon \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ with $\phi_n(a) = a$ is a surjective ring homomorphism.

b. The multiplication-by-n map $\psi_n \colon \mathbb{Z} \to \mathbb{Z}$ with $\psi_n(a) = na$ is not a ring homomorphism unless $n = 1$.

Here are several standard ring homomorphisms.

DEFINITION 3.5.3. Let R and S be rings.

a. The *identity homomorphism* $\mathrm{id}_R \colon R \to R$ is the ring homomorphism given by $\mathrm{id}_R(a) = a$ for all $a \in R$.

b. If S is a subring of R, we have the *inclusion map* $\iota_S \colon S \to R$ with $\iota_S(b) = b$ for all $b \in S$.

An inclusion map is always injective, but will only be surjective if the subring is the whole ring. Here are some other examples.

EXAMPLES 3.5.4. Let R be a nonzero ring.

a. There is an injective ring homomorphism $\iota \colon R \to R[x]$ that sends $a \in R$ to the constant polynomial $a \in R[x]$.

b. There is a surjective ring homomorphism $\pi \colon R[x] \to R$ that sends $f \in R[x]$ to its constant coefficient. Note that $\pi \circ \iota = \mathrm{id}_R$, but $\iota \circ \pi \neq \mathrm{id}_R$.

We mention another useful class of ring homomorphisms of polynomial rings, arising from maps on coefficients.

EXAMPLES 3.5.5. Let R and S be rings, and let $\phi \colon R \to S$ be a ring homomorphism. This induces maps on polynomial rings and matrix rings, as follows.

a. The map $\tilde{\phi} \colon R[x] \to S[x]$ induced by ϕ on polynomial rings is given by the formula

$$\tilde{\phi}\Big(\sum_{i=0}^{N} a_i x^i\Big) = \sum_{i=0}^{N} \phi(a_i) x^i$$

for $a_i \in R$ for $0 \le i \le N$ for some $N \ge 0$.

b. The map $\tilde{\phi} \colon M_n(R) \to M_n(S)$ induced by ϕ on matrix rings is given by the formula

$$\tilde{\phi}((a_{ij})) = (\phi(a_{ij}))$$

for $(a_{ij}) \in M_n(R)$.

REMARK 3.5.6. If R is a subring of S, then we may use the map of polynomial rings induced by the inclusion map of R into S to view $R[x]$ as a subring of $S[x]$.

REMARK 3.5.7. The product of ring homomorphisms $\phi_i \colon R_i \to S_i$ over an index set $i \in I$ is a ring homomorphism between the corresponding products.

LEMMA 3.5.8. *Let R be a ring and S be an integral domain, and let $\phi \colon R \to S$ be a nonzero homomorphism. If $u \in R^\times$, then $\phi(u) \in S^\times$.*

PROOF. Let v be a multiplicative inverse to u in R. By the previous lemma

$$\phi(u)\phi(v) = \phi(uv) = \phi(1) = 1,$$

and, similarly, we have $\phi(v)\phi(u) = \phi(vu) = 1$. □

We also have the following.

DEFINITION 3.5.9. If $R = \prod_{i \in I} R_i$ is a product of rings, then there are *projection maps*

$$\pi_i \colon R \to R_i, \qquad \pi_i((a_i)_{i \in I}) = a_i$$

which are ring homomorphisms.

REMARK 3.5.10. If $R = \prod_{i \in I} R_i$ is a product of rings, the inclusion maps $\iota_i \colon R_i \to R$ for $i \in I$ given by taking $a \in R_i$ to the element with ith coordinate a and jth coordinate 0 for $j \neq i$ are not ring homomorphisms if at least two R_i are nonzero rings, since $\iota(1) \neq 1 \in R$.

As with group homomorphisms, we have notions of kernel and image of a ring homomorphism.

DEFINITION 3.5.11. Let $\phi \colon R \to S$ be a ring homomorphism. Then the kernel of ϕ is

$$\ker \phi = \{r \in R \mid \phi(r) = 0\},$$

and the image of ϕ is

$$\operatorname{im} \phi = \{\phi(r) \mid r \in R\}.$$

One can check very easily that $\operatorname{im} \phi$ is a subring of S for any ring homomorphism $\phi \colon R \to S$. However, while $\ker \phi$ is a subgroup of R closed under multiplication, it will not contain 1 unless $\phi = 0$.

EXAMPLES 3.5.12. Let R be a ring. We consider the homomorphisms of Example 3.5.4.

a. The inclusion $\iota \colon R \to R[x]$ has $\ker \iota = 0$ and $\operatorname{im} \iota$ the subring of constant polynomials in $R[x]$, which we also denote R.

b. The projection $\pi \colon R[x] \to R$ has $\operatorname{im} \iota = R$ and kernel consisting of the polynomials with 0 constant coefficient, which is the to say, the multiples of x.

Note that since any ring homomorphism is, in particular, a homomorphism of abelian groups under addition, we have the following.

LEMMA 3.5.13. *A ring homomorphism $\phi \colon R \to S$ is injective if and only if* $\ker \phi = \{0\}$.

We will have much more to say about kernels later. For now, let us finish with a corollary for fields.

LEMMA 3.5.14. *Let $\phi \colon F \to F'$ be a nonzero ring homomorphism, where F and F' are fields. Then ϕ is injective and $\phi(x)^{-1} = \phi(x^{-1})$ for all $x \neq 0$.*

PROOF. For any $x \in F^{\times}$, we have

$$\phi(1) = \phi(x \cdot x^{-1}) = \phi(x) \cdot \phi(x^{-1}),$$

so $\phi(x)$ is nonzero and has multiplicative inverse $\phi(x^{-1})$. In particular, Lemma 3.5.13 tells us that ϕ is injective. □

As usual, we can speak about injective and surjective ring homomorphisms, as well as isomorphisms.

DEFINITION 3.5.15. A ring homomorphism $\phi \colon R \to S$ is an *isomorphism* if it is bijective.

For instance, let us check that a polynomial ring in two variables is independent of the ordering of the variables, up to an isomorphism. We leave it to the reader to treat the case of more than two variables using the following lemma and the construction in Example 1a.

LEMMA 3.5.16. *Let x and y be indeterminates. The map $\sigma\colon (R[x])[y] \to (R[y])[x]$ satisfying*

$$\sigma\left(\sum_{j=0}^{M}\left(\sum_{i=0}^{N} a_{ij}x^i\right)y^j\right) = \sum_{i=0}^{N}\left(\sum_{j=0}^{M} a_{ij}y^j\right)x^i, \tag{3.5.1}$$

where the a_{ij} are elements of R. is an isomorphism.

PROOF. Note that every element of $(R[x])[y]$ may be expressed in the form on the left of (3.5.1), since a polynomial in y with coefficients in $R[x]$ has finite degree (at most M), and each of the finitely many nonzero coefficients then has a degree, and we choose N to be at least the maximum of these degrees. Similarly, every element of $(R[y])[x]$ may be written in the form on the right of (3.5.1), so the map is onto. By definition, it is one-to-one, and we leave it to the reader to check that it is a ring homomorphism. □

As usual, the inverse of an isomorphism of rings is an isomorphism of rings.

3.6. Subrings generated by elements

DEFINITION 3.6.1. Let R be a subring of a ring S, and let X be a set of elements of S. The *subring of S generated over R by X* is the smallest subring of S containing R and X.

Since the intersection of subrings containing a given set of elements is a subring, Definition 3.6.1 makes sense. When we have a finite set X, we often speak of the subring generated over R by the elements of X, as opposed to X itself. We will only be interested in a special case in which the elements we are adding to the subring commute with every element in that subring. We note the following, which we leave to the reader to verify.

DEFINITION 3.6.2. Let R be a subring of a ring S, and let $\alpha \in S$ commute with every element of R. The ring given by adjoining α to R is

$$R[\alpha] = \left\{ \sum_{i=0}^{N} r_i\alpha^i \mid r_i \in R \text{ for all } 0 \le i \le N \text{ for some } N \ge 0 \right\}.$$

REMARK 3.6.3. We often read $R[\alpha]$ as “R adjoin α.”

We leave it to the reader to check the following.

LEMMA 3.6.4. *Let R be a subring of a ring S, and let $\alpha \in S$ commute with every element of R. The $R[\alpha]$ is the subring generated over R by α.*

DEFINITION 3.6.5. Let R be a subring of S. If $\alpha_1, \alpha_2, \dots, \alpha_n \in S$ commute with each other and every element of R, we set

$$R[\alpha_1, \alpha_2 \dots, \alpha_n] = (((R[\alpha_1])[\alpha_2])\cdots)[\alpha_n].$$

REMARK 3.6.6. The ring $R[\alpha_1, \dots, \alpha_n]$ in Definition 3.6.5 is the smallest subring of S containing R and each α_i, so generated over R by the α_i.

EXAMPLES 3.6.7.

a. The ring $\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}$ is a subring of $\mathbb{C}$ known as the Gaussian integers. Note that since $i^2 = -1$, it is unnecessary to consider polynomials of higher degree.

b. The ring

$$\mathbb{Z}[\sqrt[n]{2}] = \left\{ \sum_{i=0}^{n-1} a_i \sqrt[n]{2} \mid a_i \in \mathbb{Z}, 0 \le i \le n-1 \right\}$$

is a subring of $\mathbb{R}$.

c. The ring

$$\mathbb{Z}[i, \sqrt{2}] = \{a + bi + c\sqrt{2} + d\sqrt{2}i \mid a, b, c, d \in \mathbb{Z}\}$$

is a subring of $\mathbb{C}$.

d. The ring $\mathbb{Q}[x^2]$ is a subring of $\mathbb{Q}[x]$ consisting of polynomials of the form

$$\sum_{i=0}^{N} a_i x^{2i}$$

with every $a_i \in \mathbb{Q}$.

We may relate this to the evaluation of polynomial rings at ring elements.

DEFINITION 3.6.8. Let R be a subring of a ring S, and let $\alpha \in S$ commute with every element of R. For $f = \sum_{i=0}^{N} c_i x^i \in R[x]$, we define the value of f at $\alpha \in R$ to be

$$f(\alpha) = \sum_{i=0}^{N} c_i \alpha^i.$$

For any $a \in R$, the *evaluation-at-α map* is defined by

$$e_\alpha \colon R[x] \to R[\alpha], \qquad e_\alpha(f) = f(\alpha)$$

for all $f \in R[x]$.

The following is a result of the definitions of addition and multiplication in $R[x]$.

LEMMA 3.6.9. *Let R be a subring of a ring S, and let $\alpha \in S$ commute with every element of R. The evaluation-at-α map $e_\alpha \colon R[x] \to R[\alpha]$ is a ring homomorphism.*

PROOF. Let $f = \sum_{i=0}^{N} c_i x^i \in R[x]$, and let $g = \sum_{i=0}^{M} d_i x^i \in R[x]$ for some $n \ge 0$. Then we have

$$e_\alpha(f+g) = (f+g)(\alpha) = \sum_{i=0}^{D} (c_i + d_i)\alpha^i = \sum_{i=0}^{D} c_i \alpha^i + \sum_{i=0}^{D} d_i \alpha^i = e_\alpha(f) + e_\alpha(g),$$

where $D = \max\{M, N\}$, and

$$e_\alpha(fg) = \sum_{k=0}^{M+N} \left(\sum_{i=0}^{k} c_i d_{k-i} \right) \alpha^k.$$

Since α commutes with every element of R, we have $c_i d_{k-i}\alpha^k = c_i\alpha^i d_{k-i}\alpha^{k-i}$ for all $i \le k$, so the latter term equals

$$\left(\sum_{i=0}^{N} c_i\alpha^i\right)\cdot\left(\sum_{j=0}^{M} d_j\alpha^j\right) = e_\alpha(f)\cdot e_\alpha(g).$$

□

REMARK 3.6.10. The evaluation-at-zero map is none other than the ring homomorphism constructed in Example 3.5.4a that takes a polynomial to its constant term.

EXAMPLE 3.6.11. If X is a set and R is a ring, then the set $\mathrm{Maps}(X,R)$ of functions from X to R forms a ring under the usual operations of pointwise addition and multiplication on R. Given $a \in X$, we again have an evaluation-at-a map

$$\varepsilon_a\colon\ \mathrm{Maps}(X,R) \to R,$$

given by $\varepsilon_a(f) = f(a)$ for $f \in \mathrm{Maps}(X,R)$ and $a \in X$, which is a ring homomorphism.

EXAMPLE 3.6.12. Let R be a commutative ring, and let $a \in R$. The evaluation map e_a on $R[x]$ can be viewed as the composition $\varepsilon_a \circ \kappa$, where

$$\kappa\colon R[x] \to \mathrm{Maps}(R,R), \qquad \kappa(f)(a) = f(a)$$

for $f \in R[x]$ and $a \in R$. In other words, κ takes a polynomial to the function it defines. It is a ring homomorphism since R is commutative.

Note that even if R is commutative, κ is not always injective. For instance, if $R = \mathbb{Z}/p\mathbb{Z}$ for a prime number p, then $f = x^p - x$ is a nonzero polynomial in $R[x]$, but $p(a) = 0$ for all $a \in \mathbb{Z}/p\mathbb{Z}$, so $\kappa(p) = 0$.

3.7. Fields of fractions

As is seen by the most basic case of the integers $\mathbb{Z}$, not all rings are fields. Yet, $\mathbb{Z}$ is contained in many fields, the smallest being $\mathbb{Q}$, the rational numbers. The field $\mathbb{Q}$ consists exactly of fractions $\frac{a}{b}$, where a and b are integers and b is nonzero. One can ask more generally, given an ring R, does one have a good notion of a fraction $\frac{a}{b}$ with $a,b \in R$ and $b \neq 0$? And, if so, can one form a field out of them? As we shall, see in the case of an integral domain, the answer is yes.

LEMMA 3.7.1. *Let R be an integral domain, and set*

$$X = \{(a,b) \in R \times R \mid b \neq 0\}.$$

The relation $\sim$ on X given by $(a,b) \sim (c,d)$ if and only if $ad = bc$ is an equivalence relation.

PROOF. For $(a,b) \in X$, we have $ab = ba$, so $(a,b) \sim (a,b)$, so $\sim$ is reflexive. If $(c,d) \in X$ with $(a,b) \sim (c,d)$, then $ad = bc$ implies $cb = da$, so $(c,d) \sim (b,a)$ as well, and $\sim$ is symmetric. Finally, if $(e,f) \in X$ as well and $(a,b) \sim (c,d)$ while $(c,d) \sim (e,f)$, we have $ad = bc$ and $cf = de$. Multiplying the former equality by f and then applying the latter, we obtain

$$adf = bcf = bde.$$

Since $d \neq 0$ and R is an integral domain, this implies $af = be$, which means that $(a,b) \sim (e,f)$, and therefore $\sim$ is transitive. □

Note that the last step shows the need for having an integral domain in order to have an equivalence relation in Lemma 3.7.1.

DEFINITION 3.7.2. Let R be an integral domain. We let $Q(R)$ denote the set of equivalence classes of elements of X under the relation $\sim$ of Lemma 3.7.1. The equivalence class of (a,b) with $a,b \in R$ and $b \neq 0$ will be denoted $\frac{a}{b}$, and it is called the quotient of a by b. By using the symbol $\frac{a}{b}$, we are implicitly representing the quotient by (a,b), and this representative is called a fraction. We then refer to a as the numerator of $\frac{a}{b}$ and b as the denominator of $\frac{a}{b}$.

The following is immediate.

LEMMA 3.7.3. *Let R be an integral domain, and let $a,b,x \in R$ with b and x nonzero. Then we have*

$$\frac{a}{b} = \frac{ax}{bx}$$

in $Q(R)$.

LEMMA 3.7.4. *Let R be an integral domain. There are well-defined operations $+$ and $\cdot$ on $Q(R)$ given by*

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

and

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}.$$

PROOF. Let X be as in Lemma 3.7.1. Define $+$ on X by

$$(a,b)+(c,d) = (ad+bc,bd)$$

and $\cdot$ on X by

$$(a,b)\cdot(c,d) = (ac,bd).$$

To prove the proposition, we must show that if $(a,b) \sim (a',b')$ and $(c,d) \sim (c',d')$, we have

$$(a,b)+(c,d) \sim (a',b')+(c',d') \quad \text{and} \quad (a,b)\cdot(c,d) = (a',b')\cdot(c',d').$$

We check that

$$(ad+bc)b'd' = ab'dd' + cd'bb' = ba'dd' + dc'bb' = bd(a'd'+b'c')$$

and

$$acb'd' = ab'cd' = ba'c'd = bda'c',$$

as desired. □

COROLLARY 3.7.5. *Let R be an integral domain and $a,a',b \in R$ with $b \neq 0$. In $Q(R)$, one has*

$$\frac{a}{b} + \frac{a'}{b} = \frac{a+a'}{b}.$$

PROOF. Noting Lemma 3.7.3, we have

$$\frac{ab+a'b}{b^2} = \frac{(a+a')b}{b^2} = \frac{a+a'}{b}.$$

□

THEOREM 3.7.6. *Let R be an integral domain. Under the operations $+$ and $\cdot$ of Lemma 3.7.4, the ring $Q(R)$ is a field.*

PROOF. First, we note that addition is commutative since

$$\frac{a}{b}+\frac{c}{d}=\frac{ad+bc}{bd}=\frac{cb+da}{db}=\frac{c}{d}+\frac{a}{b},$$

and it is associative since

$$\left(\frac{a}{b}+\frac{c}{d}\right)+\frac{e}{f}=\frac{ad+bc}{bd}+\frac{e}{f}=\frac{(ad+bc)f+bde}{bdf}$$
$$=\frac{adf+b(cf+de)}{bdf}=\frac{a}{b}+\frac{cf+de}{df}=\frac{a}{b}+\left(\frac{c}{d}+\frac{e}{f}\right).$$

Next, we note that

$$\frac{0}{1}+\frac{a}{b}=\frac{a\cdot 1+0\cdot b}{1\cdot b}=\frac{a}{b},$$

so $0=\frac{0}{1}$ in $Q(R)$. We also have

$$\frac{-a}{b}+\frac{a}{b}=\frac{-ab+ab}{b^2}=\frac{0}{b^2}=\frac{0}{1},$$

the latter step by noting that $0\cdot 1=b^2\cdot 0=0$. Hence, $Q(R)$ is an abelian group under addition.

We note that multiplication is associative, as

$$\left(\frac{a}{b}\cdot\frac{c}{d}\right)\cdot\frac{e}{f}=\frac{ac}{bd}\cdot\frac{e}{f}=\frac{ace}{bdf}=\frac{a}{b}\cdot\frac{ce}{df}=\frac{a}{b}\cdot\left(\frac{c}{d}\cdot\frac{e}{f}\right).$$

We check distributivity as follows:

$$\frac{a}{b}\cdot\left(\frac{c}{d}+\frac{e}{f}\right)=\frac{a}{b}\cdot\frac{cf+de}{df}=\frac{acf+ade}{bdf}=\frac{acf}{bdf}+\frac{ade}{bdf}=\frac{ac}{bd}+\frac{ae}{bf}=\frac{a}{b}\cdot\frac{c}{d}+\frac{a}{b}\cdot\frac{e}{f}.$$

Note that

$$\frac{1}{1}\cdot\frac{a}{b}=\frac{a}{b},$$

so $1=\frac{1}{1}$ in $Q(R)$. Finally, note that $\frac{a}{b}\neq 0=\frac{0}{1}$ if and only if $a\neq 0$, and in this case we can form $\frac{b}{a}$. We then have

$$\frac{b}{a}\cdot\frac{a}{b}=\frac{ab}{ab}=\frac{1}{1}=1,$$

so $\frac{b}{a}=(\frac{a}{b})^{-1}$. Therefore, $Q(R)$ is a field. □

DEFINITION 3.7.7. Let R be an integral domain. The field $Q(R)$ is called the *quotient field*, or the *field of fractions*, of R.

REMARK 3.7.8. The field $Q(R)$ is not a quotient of R in the sense it is the set of equivalence classes for an equivalence relation on R itself. Rather, it is a set of quotients of elements of R in the sense of division, and in fact it contains R. That is, quotient rings and quotient fields are quite different should not be confused with each other.

DEFINITION 3.7.9. Let F be a field. The field $F(x)$ of fractions of $F[x]$ is called the *field of rational functions in one variable* over F.

EXAMPLE 3.7.10. The fraction $\frac{x+1}{x^2+1}$ is an element of $\mathbb{Q}(x)$, as is $\frac{x^2-x}{x^2+1}$, and

$$\frac{x+1}{x^2+1} + \frac{x^2-x}{x^2+1} = \frac{x^2+1}{x^2+1} = 1.$$

The following theorem says, in essence, that $Q(R)$ is the smallest field containing R.

THEOREM 3.7.11. *Let R be an integral domain.*

a. The map $\iota_R \colon R \to Q(R)$ given by $\iota_R(r) = \frac{r}{1}$ is an injective ring homomorphism. We use it to identify R with a subring of $Q(R)$, setting $r = \frac{r}{1}$.

b. If F is any field containing R, then is an injective ring homomorphism, then there is a unique injective homomorphism $Q(R) \to F$ that restricts to the inclusion map $R \to F$.

PROOF. That ι_R is a ring homomorphism is easily checked, and it is injective since $\frac{r}{1} = \frac{0}{1}$ implies by definition that $r = 0$. Now, suppose that R is contained a field F. Define $\theta \colon Q(R) \to F$ by

$$\theta\left(\frac{a}{b}\right) = ab^{-1}.$$

This is well-defined, as if $ad = bc$ for some $c, d \in R$ with $d \neq 0$, then $ab^{-1} = cd^{-1}$. Moreover, for any quotients $\frac{a}{b}$ and $\frac{c}{d}$ in $Q(R)$, we have

$$\theta\left(\frac{a}{b} + \frac{c}{d}\right) = (ad+bc)(bd)^{-1} = (ad+bc)\cdot(bd)^{-1} = ab^{-1} + cd^{-1} = \theta\left(\frac{a}{b}\right) + \theta\left(\frac{c}{d}\right)$$

and

$$\theta\left(\frac{a}{b}\cdot\frac{c}{d}\right) = ac \cdot bd^{-1} = ab^{-1}cd^{-1} = \theta\left(\frac{a}{b}\right)\cdot\theta\left(\frac{c}{d}\right),$$

so θ is a ring homomorphism. If $\theta(\frac{a}{b}) = 0$, then $ab^{-1} = 0$, which implies that $a = 0$, and hence $\frac{a}{b} = 0$. Therefore, θ is injective. Also, note that

$$\theta(a) = \theta\left(\frac{a}{1}\right) = a \cdot 1^{-1} = a.$$

Finally, if $\chi \colon Q(R) \to F$ is any homomorphism with which restricts to the inclusion map $R \to F$, then we have

$$\chi\left(\frac{a}{b}\right) = \chi\left(\frac{a}{1}\cdot\frac{1}{b}\right) = \chi(a)\cdot\chi(b)^{-1} = ab^{-1} = \theta\left(\frac{a}{b}\right),$$

so $\chi = \theta$. □

Let us make our comment prior to the theorem more precise.

COROLLARY 3.7.12. *Let R be an integral domain and K a field containing it. Then there is a smallest subfield F of K containing R, and it is isomorphic to the field of fractions of R via a map $Q(R) \to K$ that extends the identity map on R.*

PROOF. The smallest field F containing R is simply the intersection of all fields contained in K and containing R. We then apply Theorem 3.7.11 to the inclusion map $\iota\colon R \to F$. The image of the induced map $\theta\colon Q(R) \to F$ is a field containing R and contained in K, so must be F itself. □

Corollary 3.7.12 allows us think more concretely about fields of fractions by speaking of fields of fractions inside a given field.

DEFINITION 3.7.13. Let R be an integral domain and K a field containing it. The *field of fractions* of R in K is the smallest subfield F of K containing R.

The next corollary tells us that it's okay to think of elements of a field F of the form ab^{-1} with $a \in F$ and $b \in F^{\times}$ as fractions $\frac{a}{b}$.

COROLLARY 3.7.14. *If F is a field, then it is isomorphic to its own field of fractions.*

PROOF. By Corollary 3.7.12, there is a field containing F in F, which of course is F itself, that is isomorphic to the field of fractions $Q(F)$. □

In other words, the field of fractions of F is F.

COROLLARY 3.7.15. *Let R and S be integral domains, and let $\phi\colon R \to S$ be an injective ring homomorphism. Then there is a unique homomorphism $Q(\phi)\colon Q(R) \to Q(S)$ such that $Q(\phi)(a) = \phi(a)$ for all $a \in R$.*

PROOF. As the composite map $\iota_S \circ \phi\colon R \to Q(S)$ is injective, Theorem 3.7.11 tells us that there is a unique injective homomorphism $Q(\phi)\colon Q(R) \to Q(S)$ with $Q(\phi) \circ \iota = \iota_S \circ \phi$, as desired. □

EXAMPLE 3.7.16. The quotient field of $\mathbb{Z}[i]$ is isomorphic to $\mathbb{Q}(i)$. To see this, note that $\mathbb{Q}(i)$ is a field containing $\mathbb{Z}[i]$, and so there is an inclusion homomorphism $Q(\mathbb{Z}[i]) \to \mathbb{Q}(i)$ that takes a fraction of the form $\frac{a+bi}{c+di}$ with $a,b,c,d \in \mathbb{Z}$ and $(c,d) \neq (0,0)$ to itself, but every element in $\mathbb{Q}(i)$ has the form $q+ri$ with $q,r \in \mathbb{Q}$, and any such element can be written as such a fraction with $d = 0$.

EXAMPLE 3.7.17. The quotient field of $\mathbb{Z}[x]$ is $\mathbb{Q}(x)$, the quotient field of $\mathbb{Q}[x]$. To see this, note that the inclusion map $\alpha\colon \mathbb{Z}[x] \to \mathbb{Q}[x]$ sending a polynomial to itself induces an injective homomorphism $Q(\alpha)\colon Q(\mathbb{Z}[x]) \to \mathbb{Q}(x)$ by Corollary 3.7.15. Moreover, for $f,g \in \mathbb{Z}[x]$, we have $Q(\alpha)(\frac{f}{g}) = \frac{f}{g}$ by definition. If $f,g \in \mathbb{Q}[x]$, then there exists a nonzero $a \in \mathbb{Z}$ such that $af, ag \in \mathbb{Z}[x]$. (Here, a is the least common multiple of the denominators of the coefficients of P and Q, written as fractions in lowest terms.) Then $\frac{P}{Q} = \frac{aP}{aQ}$ in $\mathbb{Q}(x)$, so $\frac{P}{Q}$ is in the image of the map $Q(\alpha)$. Therefore, $Q(\alpha)$ is an isomorphism.

3.8. Ideals and quotient rings

In this section, we introduce the notion of an ideal of a ring. An ideal plays the role that a normal subgroup does in group theory, which is to say that we can take a quotient of a ring by an ideal and obtain another ring. The issue with simply using a subring can be seen in the following example.

EXAMPLE 3.8.1. Consider the quotient group $\mathbb{Q}/\mathbb{Z}$ under addition. The multiplication in $\mathbb{Q}$ does not induce a well-defined multiplication on $\mathbb{Z}$. To see this, note that one would like

$$(a+\mathbb{Z})\cdot(b+\mathbb{Z}) = (ab+\mathbb{Z})$$

for any $a,b \in \mathbb{Q}$. But then we would have

$$0+\mathbb{Z} = (0+\mathbb{Z})\cdot\Big(\frac{1}{2}+\mathbb{Z}\Big) = (1+\mathbb{Z})\cdot\Big(\frac{1}{2}+\mathbb{Z}\Big) = \frac{1}{2}+\mathbb{Z},$$

which is clearly not the case.

To fix this, we introduce the notion of an ideal. We begin with left and right ideals.

DEFINITION 3.8.2. A subset I of a ring R that is a subgroup under addition is called a *left* (resp., *right*) *ideal* if $R\cdot I \subseteq I$ (resp., $I\cdot R \subseteq I$).

DEFINITION 3.8.3. A two-sided ideal, or more simply, an ideal, of a ring R is any subset of R that is both a left and a right ideal.

In other words, a left ideal J of R is an additive subgroup for which $r\cdot b \in J$ for all $r \in R$ and $b \in J$, and a right ideal K is one for which $c\cdot r \in K$ for all $r \in R$ and $c \in K$. An ideal I of R is an additive subgroup for which both $r\cdot a \in I$ and $a\cdot r \in I$ for all $r \in R$ and $a \in I$.

REMARK 3.8.4. Note that $I \subseteq R\cdot I$, so the condition that $R\cdot I \subseteq I$ (resp., $I\cdot R \subseteq I$) amounts to $R\cdot I = I$ (resp., $I\cdot R = I$).

In fact, we have the following simple criterion for a nonempty subset to be an ideal.

LEMMA 3.8.5. *Let R be a ring, and let I be a nonempty subset of R. Then I is a left (resp., right ideal) if and only if the following hold:*

i. I is closed under addition: if $a,b \in I$, then $a+b \in I$, and

ii. I is closed under left (resp., right) multiplication by elements of R: if $r \in R$ and $a \in I$, then $ra \in I$ (resp., $ar \in I$).

PROOF. We need only see that a set I satisfying (i) and (ii) is a subgroup. For this, we must show that it contains 0, which it does since $0 = 0\cdot a$ for any $a \in I$, and that it contains additive inverses, which it does since $-a = -1\cdot a$ for any $a \in I$. □

REMARK 3.8.6. Every left and every right ideal in a commutative ring R is an ideal of R.

EXAMPLES 3.8.7.

a. The subset $n\mathbb{Z}$ of $\mathbb{Z}$ is an ideal of $\mathbb{Z}$ for each $n \in \mathbb{Z}$. That is, any integer multiple of an integer multiple of n is an integer multiple of n.

b. The subset $\mathbb{Z}$ of $\mathbb{Q}$ is not an ideal, as $1/2\cdot\mathbb{Z} \not\subseteq \mathbb{Z}$, for instance.

c. Let R be a nonzero ring. Consider the set of matrices in $M_n(R)$ that are 0 in all entries outside their first columns. This is a left ideal of $M_n(R)$, but it is not a right ideal for $n \geq 2$. Similarly, the set of matrices in $M_n(R)$ that are 0 in all entries outside their first rows is a right ideal of $M_n(R)$.

d. Let R be a ring. The set of all polynomials with zero constant coefficient is an ideal of R, equal to the set of multiples of x in $R[x]$.

DEFINITION 3.8.8. The zero ideal of a ring R is the subset $\{0\}$. The improper ideal of R is the ring R itself. An ideal is said to be nonzero if it is not equal to zero, and an ideal is said to be proper if it is not equal to R.

We note the following.

LEMMA 3.8.9. *Let R be a ring, and let I be a left (or right) ideal of R. Then $I = R$ if and only if I contains a unit, and in particular if and only if I contains* 1.

PROOF. If $I = R$, then clearly I contains 1 and therefore a unit. If $u \in I$ is a unit, then $u^{-1} \in R$, so $1 = u^{-1} \cdot u \in I$. And if $1 \in I$, then $a = a \cdot 1 \in I$ for all $a \in R$. □

The following classifies, as a special case, all ideals in a field.

COROLLARY 3.8.10. *The only left and only right ideals in a division ring are $\{0\}$ and D.*

PROOF. If I is a nonzero left or right ideal of D, it then contains a unit, so is D. □

We shall see later that the converse to Corollary 3.8.10 also holds. We give one more example.

LEMMA 3.8.11. *Let R and S be rings with unity. Then any left ideal of $R \times S$ has the form $I \times J$, where I is a left ideal of R and J is a left ideal of S.*

PROOF. Let K be an ideal of $R \times S$. Let

$$I = \{a \in R \mid (a,0) \in K\} \quad \text{and} \quad J = \{b \in S \mid (0,b) \in K\},$$

which are left ideals of R and of S, respectively. If $(a,b) \in K$, then $(1,0) \cdot (a,b) = (a,0)$, so $a \in I$ and $(1,0) \cdot (a,b) = (0,b)$, so $b \in J$. Therefore, $K \subseteq I \times J$. Conversely, if $(a,b) \in I \times J$, then $(a,0) \in K$ and $(0,b) \in K$, so $(a,b) = (a,0) + (0,b) \in K$, so $I \times J \subseteq K$. □

The following is the ring-theoretic analogue of Proposition 2.12.11.

PROPOSITION 3.8.12. *Let $\phi \colon R \to S$ be a homomorphism of rings. Then* $\ker\phi$ *is an ideal of R.*

PROOF. We know from Proposition 2.8.8 that $\ker\phi$ is a subgroup of R under addition. Moreover, if $r \in R$ and $a \in \ker\phi$, then

$$\phi(ra) = \phi(r)\phi(a) = \phi(r) \cdot 0 = 0,$$

so $ra \in \ker\phi$. Similarly, we have $\phi(ar) = 0$, so $ar \in \ker\phi$ as well. □

We may now construct the analogue of a quotient group, known as a quotient ring.

THEOREM 3.8.13. *Let R be a ring, and let I be a two-sided ideal of R. Then the quotient group R/I has a well-defined multiplication on it, given by*

$$(r+I) \cdot (s+I) = (rs+I)$$

for $r,s \in R$. Moreover, with the usual addition of cosets and this multiplication, R/I becomes a ring.

PROOF. Suppose that $a,a',b,b' \in R$ with $a+I = a'+I$ and $b+I = b'+I$. Then there exist $x,y \in I$ with $a' = a+x$ and $b' = b+y$. We have

$$(a'b'+I) = (a+x)(b+y)+I = ab+ay+xb+xy+I = ab+I,$$

since $ay, xb, xy \in I$ in that I is a two-sided ideal. Therefore, the multiplication on R/I is well-defined. That it is associative is a direct consequence of the associativity of multiplication on R. Distributivity is again a consequence of distributivity on R, but we write out the proof of the left distributive law:

$$\begin{aligned}(a+I)\cdot((b+I)+(c+I)) = (a+I)\cdot(b+c+I) = a(b+c)+I &= (ab+ac)+I \\ = (ab+I)+(ac+I) = (a+I)\cdot(b+I) &+ (a+I)\cdot(c+I).\end{aligned}$$

□

DEFINITION 3.8.14. The quotient of a ring R by an ideal I is the ring R/I defined by Theorem 3.8.13. We say that R/I is the quotient ring of R by I (or the factor ring of R by I).

EXAMPLES 3.8.15.

a. The quotient of the ring $\mathbb{Z}$ by the ideal $n\mathbb{Z}$ is the ring $\mathbb{Z}/n\mathbb{Z}$.

b. The quotient of any ring R by the zero ideal is isomorphic to R. The quotient of any ring R by R is isomorphic to the zero ring.

The following is immediately verified.

DEFINITION 3.8.16. The map $\pi_I \colon R \to R/I$ defined by $\pi_I(a) = a+I$ is called the quotient map from R to R/I.

REMARK 3.8.17. For R a ring, I and ideal of R, and $a,b \in R$, we may sometimes write $a \equiv b \bmod I$ to mean that $a+I = b+I$, or simply just $a = b$ when it is understood that we are working with the images of a and b under π_I, i.e., in the ring R/I.

The following is easily verified.

LEMMA 3.8.18. *Let I be an ideal in a ring R. The quotient map $\pi_I \colon R \to R/I$ is a surjective ring homomorphism with kernel I.*

We have the analogue of the first isomorphism theorem.

THEOREM 3.8.19. *Let $\phi \colon R \to S$ be a homomorphism of rings. Then the map*

$$\bar{\phi} \colon R/\ker\phi \to \operatorname{im}\phi$$

defined by $\bar{\phi}(a+\ker\phi) = \phi(a)$ for all $a \in R$ is an isomorphism of rings.

PROOF. We know that $\bar{\phi}$ is an isomorphism of additive groups by Theorem 2.13.11. Let $I = \ker\phi$. For $a,b \in R$, we have

$$\bar{\phi}((a+I)(b+I)) = \bar{\phi}(ab+I) = \phi(ab) = \phi(a)\phi(b) = \bar{\phi}(a+I)\bar{\phi}(b+I),$$

so $\bar{\phi}$ is a ring homomorphism as well, therefore, a ring isomorphism. □

EXAMPLE 3.8.20. The kernel of the homomorphism $\pi\colon R[x] \to R$ of Example 3.5.4a is the ideal I of polynomials with zero constant coefficient. In that it is onto, e_0 induces an isomorphism between $R[x]/I$ and R.

Note that if ϕ is a surjective map, then ϕ carries ideals to ideals.

PROPOSITION 3.8.21. *Let $\phi\colon R \to S$ be a surjective homomorphism of rings. If I is a left (resp., right) ideal of R, then $\phi(I)$ is a left (resp., right) ideal of S.*

PROOF. We show this for left ideals I of R. Let $s \in S$. Then $s = \phi(r)$ for some $r \in R$, and if $a \in I$, then $s\phi(a) = \phi(ra) \in \phi(I)$, so $\phi(I)$ is a left ideal of S. □

We also have the following generalization of Proposition 3.8.12, the proof of which we leave to the reader.

PROPOSITION 3.8.22. *Let $\phi\colon R \to S$ be a ring homomorphism, and let J be a left (resp., right) ideal of S. Then $\phi^{-1}(J)$ is a left (resp., right) ideal of R.*

We can now classify the ideals in quotient rings.

THEOREM 3.8.23. *Let R be a ring, and let I be an ideal of R. Then the quotient map $\pi_I\colon R \to R/I$ induces a one-to-one correspondence between the left, right, and two-sided ideals of R containing I and the left, right, and two-sided ideals of R/I, respectively.*

PROOF. We prove this for left ideals. If J is a left ideal of R containing I, then $\pi_I(J)$ is a left ideal of R/I by Proposition 3.8.21. If $\pi_I(J) = \pi_I(K)$ for some left ideal of R containing K, then any $j \in J$ satisfies $j = k + i$ for some $k \in K$ and $i \in I$, and therefore $j \in K$ since $I \subseteq K$. We therefore have $J \subseteq K$, and similarly $K \subseteq J$, so $J = K$. On the other hand, if N is any left ideal of R/I, then $J = \pi^{-1}(N)$ is a left ideal of R, and it contains I since $I = \pi^{-1}(\{0\})$. Since $\pi_I(J) = N$, we are done. □

3.9. Principal ideals and generators

DEFINITION 3.9.1. A left ideal J of a ring R with unity is said to be *principal* if there exists an element $a \in R$ such that

$$J = Ra = \{ra \mid r \in R\}.$$

Similarly, a right ideal K of a ring R is *principal* if there exists an element $a \in R$ such that $K = aR = \{ar \mid r \in R\}$. We then say that J (resp., K) is the left (resp, right) ideal generated by a.

REMARK 3.9.2. Note that Ra for $a \in R$ is always a left ideal of R, since $ra - sa = (r-s)a$ for $r, s \in R$, so Ra is an additive subgroup, and $s(ra) = (sr)a$, so Ra is closed under left multiplication by elements of R.

We also have the notion of a principal ideal.

DEFINITION 3.9.3. An ideal I of a ring R with unity is *principal* if there exists an element $a \in R$ such that

$$I = \left\{ \sum_{i=1}^{N} r_i a s_i \mid r_i, s_i \in R \text{ for } 1 \le i \le N \text{ and } N \ge 0 \right\}.$$

We then say that I is generated by a and write $I = (a)$.

REMARK 3.9.4. The set $RaR = \{ras \mid r,s \in R\}$ will not in general be a two-sided ideal, as $ras + r'as'$ for $r,r',s,s' \in R$ need not itself be an element of RaR.

EXAMPLES 3.9.5.

a. For each $n \geq 1$, the ideal $n\mathbb{Z}$ is the principal ideal (n).

b. For every ring R, the zero ideal is the principal ideal (0).

c. For every ring R with unity, we have $R = (1)$, so R is a principal ideal of R, known as the imp

d. The ideal (x) in $R[x]$ is the ideal consisting of all polynomials with nonconstant coefficient.

e. The ideal generated by $(a,b) \in \mathbb{Z} \times \mathbb{Z}$ is equal to the set $\{(ax,by) \mid x,y \in \mathbb{Z}\}$.

EXAMPLE 3.9.6. Let R be a nonzero ring, and let $n \geq 1$. For integers s,t with $1 \leq s,t \leq n$, let $E_{st} = (e_{ij}) \in M_n(R)$ be the matrix with $e_{st} = 1$ and $e_{ij} = 0$ for $(i,j) \neq (s,t)$. If $A = (a_{ij}) \in M_n(R)$ is any matrix, then the (i,j)th entry of AE_{ss} is a_{is} if $j = s$ and 0 otherwise. Therefore, the left ideal generated by E_{ss} is

$$M_n(R)E_{ss} = \{(b_{ij}) \in M_n(R) \mid b_{ij} = 0 \text{ for } 1 \leq i,j \leq N, j \neq s\},$$

the set of matrices that are zero outside of the sth column. Similarly, the (i,j)th entry of $E_{ss}A$ is a_{sj} if $i = s$ and 0 otherwise, so $E_{ss}M_n(R)$ is the right ideal of matrices that are zero outside of the sth row.

The two-sided ideal (E_{ss}) of $M_n(R)$ is in fact all of $M_n(R)$. To see this, note that $E_{js}E_{ss}E_{sj} = E_{jj}$ for any $j \in \mathbb{Z}$. We then have

$$A = \sum_{j=1}^{n} AE_{jj} = \sum_{j=1}^{n} (AE_{js})E_{ss}E_{sj} \in (E_{ss}).$$

Note, however, that the set $X = M_n(R)E_{ss}M_n(R)$ is not $M_n(R)$, since each column of a matrix in X has entries which are all equal to each other.

DEFINITION 3.9.7. A nonzero ring R is *simple* if its only ideals are 0 and R.

REMARK 3.9.8. The reader can check using Example 3.9.6 that if D is a division ring and $A \in M_n(D)$ is nonzero, then the ideal (A) is all of $M_n(D)$. So, $M_n(D)$ is simple, but note that it is not a division ring if $n \geq 2$, and it does have proper, nonzero left ideals.

The following three results also clearly have analogues for right ideals that we leave unstated.

PROPOSITION 3.9.9. *Let D be a ring that contains no nonzero, proper left ideals. Then D is a division ring.*

PROOF. Let $u \in D$ be nonzero. By assumption, we have $Du = D$, so there exists $v \in D$ such that $vu = 1$. Then $Dv = D$, so there exists $w \in D$ such that $wv = 1$. We then have $w = wvu = u$, so $u = v^{-1} \in D^{\times}$. □

LEMMA 3.9.10. *Let R be a ring, and let $a,b \in R$. Then $Ra \subseteq Rb$ if and only if there exists $r \in R$ such that $a = rb$.*

PROOF. If $Ra \subseteq Rb$, then since $a \in Rb$, we have $a = rb$ for some $r \in R$. Conversely, if $a = rb$ and $r' \in R$, then $r'a = (r'r)b \in Rb$, so $Ra \subseteq Rb$. □

LEMMA 3.9.11. *Let R be a ring that has no zero divisors. Let $a, b \in R$. Then $Ra = Rb$ if and only if $b = ua$ for some $u \in R^{\times}$.*

PROOF. Note that $a = 0$ if and only if $b = 0$, so we may suppose that a and b are nonzero with $Ra = Rb$. Since $b \in Ra$, we have that there exists $u \in R$ with $b = ua$. Similarly, there exists $v \in R$ with $a = vb$. But then $a = vua$ and $b = uvb$, so $(1 - vu)a = (1 - uv)b$. Since R has no left zero divisors, we have $uv = vu = 1$. Conversely, if $b = ua$, then clearly $b \in Ra$, so $Rb \subseteq Ra$. On the other hand, $a = u^{-1}b$, so $Ra \subseteq Rb$ as well. □

EXAMPLE 3.9.12. In $\mathbb{Q}[x]$, we have $(f) = (g)$ if and only if $f = cg$ for some $c \in \mathbb{Q}^{\times}$, since $\mathbb{Q}[x]^{\times} = \mathbb{Q}^{\times}$.

We have various operations that can be performed on ideals.

LEMMA 3.9.13. *Let I and J be left ideals (resp., right ideals) of a ring R.*

a. The set

$$I + J = \{a + b \mid a \in I, b \in J\}$$

is a left ideal (resp., right ideal) of R.

b. The intersection $I \cap J$ is a left (resp., right ideal) of R.

PROOF.

a. If $a \in I$, $b \in J$, and $r \in R$, then $r(a+b) = ra + rb$, and $ra \in I$, $rb \in J$ since I and J are ideals, so $r(a+b) \in I + J$. Moreover, $I + J$ is a subgroup of R under addition by Lemma 4.1.4.

b. If $a, b \in I \cap J$ and $r \in R$, then clearly $a - b \in I \cap J$ and $ra \in I \cap J$, so $I \cap J$ is a left ideal of R.

□

REMARK 3.9.14. The argument of Lemma 3.9.13b carries over to show that an arbitrary intersection of left (resp., right) ideals of a ring R is a left (resp., right) ideal of R.

Clearly, addition of ideals forms an associative and commutative binary operation on the set of ideals of a ring. More generally, we have the following result.

LEMMA 3.9.15. *Let T be an indexing set, and let $\{I_t \mid t \in T\}$ be a collection of left (resp., right ideals) of a ring R. Then the set*

$$\sum_{t \in T} I_t = \left\{ \sum_{i=1}^{N} a_{t_i} \mid t_i \in T, a_{t_i} \in I_{t_i} \text{ for each } 1 \le i \le N \text{ for some } N \ge 0 \right\}$$

of finite sums of elements of the ideals I_t is an ideal of R, equal to the intersection of all ideals of R containing I_t for every $t \in T$.

PROOF. Note that $\sum_{t\in T} I_t$ consists exactly of finite sums of elements in the union $\cup_{t\in T} I_t$. It is a subgroup, as the sum of two finite sums is a finite sum, and the negative of two finite sums is as well. Moreover, it is an ideal, as for any $N \geq 0$, $t_i \in T$ and $a_{t_i} \in I_{t_i}$ for $1 \leq i \leq N$, we have

$$r \cdot \sum_{i=1}^{N} a_{t_i} = \sum_{i=1}^{N} r a_{t_i},$$

and $ra_{t_i} \in I_{t_i}$ since I_{t_i} is a left ideal of R. Therefore $\sum_{t\in T} I_t$ is a left ideal, and similarly, it is a right ideal.

Finally, note that if J is any ideal of R containing each I_t, then it must contain any finite sum of elements in these ideals, i.e., in $\cup_{t\in T} I_t$. Therefore, J contains $\sum_{t\in T} I_t$. Therefore, the intersection of all ideals of R containing each I_t is an ideal of R containing $\sum_{t\in I_t} I_t$, and $\sum_{t\in I_t} I_t$ is itself an ideal of R containing each I_t, so it equals the intersection. □

DEFINITION 3.9.16. Let R be a ring, and let $\{I_t \mid t \in T\}$ be a collection of left (resp., right) ideals. The sum of the ideals I_t with $t \in T$ is the left (resp., right) ideal $\sum_{t\in T} I_t$ of R.

We will define generators solely for two-sided ideals, though they have obvious analogues for left and right ideals.

DEFINITION 3.9.17. Let X be a subset of a ring R with unity. The ideal (X) generated by X is the sum of the ideals (x) for $x \in X$. If I is an ideal of R and $I = (X)$, we say that X is a set of generators of I, and X generates I. The elements of X are called generators. If $X = \{a_1, a_2, \ldots, a_n\}$ is a finite set, then we write $(a_1, a_2, \ldots, a_n)$ for (X).

REMARKS 3.9.18.

a. Every ideal is generated by the set of all of its elements.

b. We could equivalently have defined (X) to be the smallest ideal containing X using Lemma 1.2.24.

Since the set-theoretic product of two ideals will not in general be closed under addition, we depart from earlier notation to make the following definition.

DEFINITION 3.9.19. Let I and J be ideals of a ring R. Then the product IJ of I and J is the ideal of R generated by all ab with $a \in I$ and $b \in J$.

In particular, we may speak of powers $I^n = II \cdots I$ of an ideal I for any $n \geq 1$. Products are easily calculated in terms of generators, as seen in the following examples.

EXAMPLES 3.9.20.

a. If R is a ring and $x, y \in R$, then $(x) \cdot (y) = (xy)$.

b. In the ring $\mathbb{Q}[x,y]$, we have

$$(x,y) \cdot (x^2, x+y) = (x^3, x^2y, x^2+xy, xy+y^2).$$

DEFINITION 3.9.21. We say that two ideals I and J of a ring R are *coprime* if $I + J = R$.

DEFINITION 3.9.22. For $k \geq 1$, we say that ideals $I_1, \ldots, I_k$ of a ring R are *pairwise coprime* if $I_i + I_j = R$ for all $1 \leq i < j \leq k$.

We prove a general form of the Chinese Remainder Theorem.

THEOREM 3.9.23 (Chinese Remainder Theorem). *Let $I_1, \dots, I_k$ be pairwise coprime two-sided ideals of a ring R for some $k \geq 1$. Then there is an isomorphism*

$$R/(I_1 \cap I_2 \cap \cdots \cap I_k) \xrightarrow{\sim} R/I_1 \times R/I_2 \times \cdots \times R/I_k$$

that sends the coset of $a \in R$ to $(a + I_1, a + I_2, \dots, a + I_k)$.

PROOF. The kernel of the map $R \to \prod_{i=1}^{k} R/I_i$ induced by the diagonal map is clearly $I_1 \cap I_2 \cap \cdots \cap I_k$. We need only see that it is surjective. Consider the case that $k = 2$. Let $a, b \in R$. Then there exist $d \in I_1$ and $c \in I_2$ such that $a + I_1 = c + I_1$ and $b + I_2 = d + I_2$. If we set $x = c + d$, then $x + I_1 = c + I_1$ and $x + I_2 = d + I_2$, so x maps to $(a + I_1, b + I_2)$.

For any $k \geq 3$, suppose by induction we know the result for $k - 1$, so $R/(I_2 \cap \cdots \cap I_k) \cong R/I_2 \times \cdots R/I_k$. We therefore need only see that I_1 and $I_2 \cap \cdots \cap I_k$ are coprime. Note that $I_2 \cap \cdots \cap I_k$ contains the product $I_2 \cdots I_k$. For each $2 \leq i \leq k$, let $a_i \in I_1$ and $b_i \in I_i$ be such that $a_i + b_i = 1$. Then $1 = (a_2 + b_2) \cdots (a_k + b_k)$ is an element of I_1 plus $b_1 \dots b_k \in I_2 \cdots I_k$, as needed. □

DEFINITION 3.9.24. An ideal I of a ring R with unity is said to be *finitely generated* if it has a finite set of generators, which is to say that $I = (a_1, a_2, \dots, a_n)$ for some $n \geq 1$ and $a_1, a_2, \dots, a_n \in I$.

EXAMPLE 3.9.25. If $R[x, y]$, the ideal (x, y) is the ideal of elements with 0 constant term, as every monomial other than 1 is either divisible x or y. It is not principal, since no element of $R[x, y]$ not in $R^{\times}$ divides both x and y, but it is finitely generated.

EXAMPLE 3.9.26. Let $n \geq 2$. The ideal (n, x) of $\mathbb{Z}[x]$ is the set of all sums $nf + xg$ with $f, g \in \mathbb{Z}[x]$, which is equal to the set of polynomials with $\mathbb{Z}$-coefficients and constant coefficient divisible by n. This is not principal, since n and x are both multiples only of ± 1, which are not contained in (n, x)

EXAMPLE 3.9.27. Consider the ideal $(4, 6)$ of $\mathbb{Z}$. It contains $2 = 6 - 4$, so $(2) \subseteq (4, 6)$ and we have $4, 6 \in (2)$, so $(4, 6) \subseteq (2)$. Therefore, $(4, 6)$ is a principal ideal of $\mathbb{Z}$, equal to the ideal (2).

In fact, note the following.

LEMMA 3.9.28. *The ideals of $\mathbb{Z}$ are exactly the subgroups of $\mathbb{Z}$ under addition, i.e., the $n\mathbb{Z}$ with $n \geq 0$. In particular, every ideal of $\mathbb{Z}$ is principal.*

PROOF. Ideals are by definition subgroups under addition, and if I is an ideal of $\mathbb{Z}$, the condition that $\mathbb{Z} \cdot I \subseteq I$ is a consequence of this, since it merely says that $\mathbb{Z}$-multiples of elements of I are contained in I. That the subgroups of $\mathbb{Z}$ have the form $n\mathbb{Z}$ is Corollary 2.3.12. □

This leads to the following definition.

DEFINITION 3.9.29. An integral domain R is a *principal ideal domain*, or *PID*, if every ideal in R is principal.

So far, we have the following examples.

EXAMPLES 3.9.30.

a. The ring $\mathbb{Z}$ is a principal ideal domain.

b. Every field is a principal ideal domain.

c. If R and S are principal ideal domains, then every ideal $R \times S$ is principal, though it is not a domain.

3.10. Polynomial rings over fields

We now focus on polynomial rings over a field. One of the key properties of polynomials with coefficients in a field is that we can divide them. The following is a the division algorithm in these rings.

THEOREM 3.10.1 (Division algorithm). *Let F be a field. Suppose that $f, g \in F[x]$ are polynomials with $g \neq 0$. Then there exist unique polynomials $q, r \in F[x]$ such that $f = qg + r$ and $\deg r < \deg g$.*

PROOF. The case that $f = 0$ is trivial, so we assume that f is nonzero. We verify this by induction on the degree n of f. Note that if $n \le \deg g$, and in particular if $n = 0$, then we may take $q = 0$ and $r = f$ if $\deg g > 0$ and $q = fg^{-1}$ and $r = 0$ if $\deg g = 0$ (recalling that we consider the degree of 0 to be less than that of every nonzero polynomial). So suppose that $n \ge m = \deg g$. Let a_n be the nonzero coefficient of x^n in f and b_m be the nonzero coefficient of x^m in g. Then $f' = f - a_n b_m^{-1} x^{n-m} g$ has degree at most n, and the coefficient of x^n is $a_n - a_n b_m^{-1} \cdot b_m = 0$, so in fact we have $\deg f' < n$. By induction, therefore, there exist q' and r in $R[x]$ such that $f' = q'g + r$ and $\deg r < m$. Setting $q = a_n b_m^{-1} + q'$, we have

$$f = a_n b_m^{-1} g + f' = (a_n b_m^{-1} + q')g + r = qg + r,$$

as desired.

If $qg + r = q'g + r'$ for some $q', r' \in F[x]$ with $\deg r' < \deg g$, then we have

$$(q - q')g + (r - r') = 0. \tag{3.10.1}$$

If $q \neq q'$, we would have

$$\deg(q - q')g \ge \deg g > \deg(r - r'),$$

in contradiction to (3.10.1). So, we must have $q = q'$, and then (3.10.1) yields $r = r'$, establishing uniqueness. □

We next show that polynomial rings in one variable over a field form another class of principal ideal domains.

THEOREM 3.10.2. *Let F be a field. Then $F[x]$ is a principal ideal domain. In fact, any nonzero ideal I of F is generated by any nonzero polynomial that has minimal degree among all polynomials in I.*

PROOF. By Theorem 3.4.13, $F[x]$ is an integral domain. Let I be a nonzero ideal in $F[x]$, and let g be a nonzero polynomial in $F[x]$ of minimal degree. We claim that $I = (g)$. Let $f \in I$. Using the division algorithm, we write $f = qg + r$ with $q, r \in F[x]$ with $\deg r < \deg g$. Then $r = f - qg \in I$, which by the minimality of the degree of g forces $r = 0$. Thus $f \in (g)$, and as f was arbitrary, we have $I = (g)$. □

DEFINITION 3.10.3. Let F be a field. A nonconstant polynomial $f \in F[x]$ is irreducible if there does not exist any $g \in F[x]$ with $0 < \deg g < \deg f$ that divides f. A nonconstant polynomial that is not irreducible is called reducible. A noncontant divisor of a polynomial is referred to as a factor.

EXAMPLE 3.10.4. By definition, any polynomial of degree 1 is irreducible in $F[x]$. The polynomial $x^2 + 1$ is irreducible in $\mathbb{Q}[x]$ but not in $\mathbb{C}[x]$, where we have

$$x^2 + 1 = (x + i)(x - i).$$

On the other hand, x^2 is reducible for any F, since $x^2 = x \cdot x$.

DEFINITION 3.10.5. Let R be a ring. We say that $a \in R$ is a root (or zero) of a polynomial $f \in R[x]$ if $f(a) = 0$.

DEFINITION 3.10.6. In a commutative ring R, we say that an element b divides an element a in R if there exists some $c \in R$ such that $a = bc$. Equivalently, b divides a if $a \in (b)$. We sometimes write $b \mid a$ to denote that b divides a.

We note the following.

PROPOSITION 3.10.7. *Let F be a field, and let $f \in F[x]$. Then $a \in F$ is a root of f if and only if $x - a$ divides f.*

PROOF. If $x - a$ divides f, then there exists $g \in F[x]$ with $f = (x - a)g$. We then have $f(a) = (a - a)g(a) = 0$, noting Lemma 3.6.9. Conversely, if a is a zero of f, then the division algorithm implies that there exists some $q \in F[x]$ and $c \in F$ such that $f = q(x - a) + c$. We then have

$$0 = f(a) = q(a)(a - a) + c = c,$$

so $x - a$ divides f. □

We obtain the following corollaries.

COROLLARY 3.10.8. *Let F be a field and $f \in F[x]$ be a polynomial of degree greater than* 1. *If f has a root in F, then f is reducible.*

PROOF. If $a \in F$ is a root of f, then Proposition 3.10.7 implies that $f = g(x - a)$ for some $g \in F[x]$ with $\deg g = \deg f - 1 > 0$, so f is not irreducible. □

Since a reducible polynomial of degree 2 or 3 must have a linear factor, we therefore have the following.

COROLLARY 3.10.9. *Let F be a field and $f \in F[x]$ be a polynomial of degree* 2 *or* 3. *Then f is reducible if and only if it has a root in F.*

COROLLARY 3.10.10. *Let F be a field, and let $f \in F[x]$ be a nonzero polynomial. Then f has at most* $\deg f$ *distinct roots in F.*

PROOF. Suppose that $f = (x-a_1)\dots(x-a_m)g$, where $g \in F[x]$ has no roots, and $a_1, a_2, \dots, a_m \in F$. Clearly, we may write f in this form, as otherwise we can factor out from g a linear term $x-b$ for some β with $g(b) = 0$. Moreover, we must have $m \leq n$ by degree considerations. Finally, if $f(c) = 0$ for some $c \in F$, then since F is an integral domain, we must have $c - a_i = 0$ for some i, which is to say that the a_i are the only roots of f. □

EXAMPLES 3.10.11.

a. The polynomial x^2 has 0 as its only root.

b. The polynomial x^2+1 has no roots in $\mathbb{Q}$, but it has two roots, ± 1, in $\mathbb{C}$.

c. The polynomial

$$x^4 - x^2 - 2x - 1 = (x^2+x+1)(x^2-x-1)$$

is not irreducible in $\mathbb{Q}[x]$, but it has no roots in $\mathbb{Q}$.

3.11. Maximal and prime ideals

Recall that $\mathbb{Z}/n\mathbb{Z}$ is a field for n prime, but $\mathbb{Z}/n\mathbb{Z}$ is not a field for n composite. In this section, we shall see how we can interpret this as a property of the ideal $n\mathbb{Z}$.

DEFINITION 3.11.1. An ideal $\mathfrak{m}$ of a ring R is *maximal* if it is a proper ideal of R that is not properly contained in any proper ideal of R.

In other words, a proper ideal $\mathfrak{m}$ of R is maximal if there does not exist an ideal N of R such that $\mathfrak{m} \subsetneq N \subsetneq R$.

EXAMPLES 3.11.2.

a. The maximal ideals of $\mathbb{Z}$ are exactly the $p\mathbb{Z}$ for p prime, as $m\mathbb{Z}$ contains $n\mathbb{Z}$ if and only if m divides n. In particular, as p is a prime number, $p\mathbb{Z}$ is not contained in $n\mathbb{Z}$ for any $n \geq 2$ with $n \neq p$.

b. In a field, the unique maximal ideal is (0).

c. In $\mathbb{Z} \times \mathbb{Z}$, the maximal ideals have either the form $p\mathbb{Z} \times \mathbb{Z}$ or $\mathbb{Z} \times p\mathbb{Z}$ for some prime number p.

PROPOSITION 3.11.3. *Let F be a field. The maximal ideals of $F[x]$ are exactly the ideals of the form (f) with $f \in F[x]$ irreducible.*

PROOF. Let $f \in F[x]$. If $f = 0$, then $(f) = 0$, which is not maximal. If f is a nonzero constant, then $(f) = (1) = F[x]$. If f is reducible, then $f = gh$ with $g,h \in F[x]$ nonconstant, and then $(f) \subseteq (g)$, but $g \notin (f)$ since $\deg g < \deg f$, so (f) is not maximal.

If f is irreducible and I is an ideal containing (f), then $I = (g)$for some $g \in F[x]$ as $F[x]$ is a PID. There then exists $h \in F[x]$ such that $f = gh$. Since f is irreducible, we then have that either g or h is constant, which is to say that $I = (g) = F[x]$ or $I = (g) = (f)$. In other words, (f) is maximal. □

The following gives an alternate characterization of maximal ideals of rings.

THEOREM 3.11.4. *A proper ideal $\mathfrak{m}$ in a commutative ring R with unity is maximal if and only if $R/\mathfrak{m}$ is a field.*

PROOF. By Theorem 3.8.23, the ideals in $R/\mathfrak{m}$ are in one-to-one correspondence with the ideals in R containing $\mathfrak{m}$, which are just $\mathfrak{m}$ and R. Since $R/\mathfrak{m}$ has just two ideals, they must be 0 and $R/\mathfrak{m}$. Therefore, every nonzero element of $R/\mathfrak{m}$ generates the ideal $R/\mathfrak{m}$, so is a unit. It follows that $R/\mathfrak{m}$ is a field. □

REMARK 3.11.5. The same argument can be applied to noncommutative rings R to conclude that if $\mathfrak{m}$ is maximal then $R/\mathfrak{m}$ has no nonzero proper ideals. However, as we have remarked above, this does not imply that $R/\mathfrak{m}$ is a division ring.

EXAMPLE 3.11.6. Recall that $\mathbb{Z}/n\mathbb{Z}$ is a field if and only if $n \geq 2$ is prime, which is to say if and only if $n\mathbb{Z}$ is a maximal ideal of $\mathbb{Z}$.

EXAMPLE 3.11.7. Since x^2+1 is irreducible over $\mathbb{Q}$, the ideal $\mathfrak{m} = (x^2+1)$ is maximal in $\mathbb{Q}[x]$. Clearly, $\mathfrak{m}$ is contained in the kernel of the evaluation map $e_i \colon \mathbb{Q}[x] \to \mathbb{Q}(i)$ defined by $e_i(f) = f(i)$, but then it must be the entire kernel as the kernel is proper and $\mathfrak{m}$ is maximal. By the first isomorphism theorem for rings, the field $\mathbb{Q}[x]/(x^2+1)$ is isomorphic to $\mathbb{Q}[i]$. In particular, $\mathbb{Q}[i]$ is equal to the subfield $\mathbb{Q}(i)$ of $\mathbb{C}$ consisting of fractions $\frac{a+bi}{c+di}$ with $a,b,c,d \in \mathbb{Q}$ and $(c,d) \neq (0,0)$. One can also see this directly: the multiplicative inverse of $c+di$ is $\frac{c}{c^2+d^2} - \frac{d}{c^2+d^2}i$.

EXAMPLE 3.11.8. The ring $\mathbb{Q}[x]/(x^2)$ is not a field, or even an integral domain, since $x \cdot x \in (x^2)$.

EXAMPLE 3.11.9. In $\mathbb{Z}[x]$, the ideals (p,x), where p is a prime number, are maximal. To see this, consider the homomorphism

$$\phi \colon \mathbb{Z}[x] \to \mathbb{Z}/p\mathbb{Z}$$

given by $\phi(f) = f(0) + p\mathbb{Z}$. This is surjective with kernel consisting of those f with constant coefficient a multiple of p, which is to say the ideal (p,x).

Given a proper ideal I of a ring R: is I necessarily even contained in a maximal ideal? Assuming the axiom of choice, the answer is yes. We require a preliminary lemma.

LEMMA 3.11.10. *Let $\mathscr{C}$ be a chain of ideals in a ring R, ordered with respect to inclusion of subsets of R. Then the ideal*

$$N = \bigcup_{J \in \mathscr{C}} J$$

is an ideal of R.

PROOF. If $x,y \in N$, then $x \in J$ and $y \in K$ for some $J,K \in \mathscr{C}$. Then $J \cup K$ is either J or K, so is in $\mathscr{C}$, and we then have $x-y \in J \cup K$, so $x-y \in N$. Thus, N is a subgroup of R under addition. For $a \in R$ and $x \in N$, we have that $x \in J$ for some $J \in \mathscr{C}$, and then ax and xa are elements of J, since J is an ideal. In particular, they are also elements of N. Therefore, N is an ideal. □

THEOREM 3.11.11. *Let I be a proper ideal of a ring R with unity. Then there exists a maximal ideal $\mathfrak{m}$ of R that contains I.*

PROOF. Let X be the set of proper ideals of R containing I, which we endow with the usual partial ordering $\subseteq$. Suppose that $\mathscr{C} \subset X$ is a chain. Consider the ideal

$$N = \bigcup_{J \in \mathscr{C}} J$$

of R. Note that $1 \notin N$ since $1 \notin J$ for all $J \in \mathscr{C}$, so N is proper. In other words, $N \in X$, and it is an upper bound for $\mathscr{C}$. Zorn's lemma then tells us that X contains a maximal element, which is necessarily a maximal ideal of R. □

In commutative rings with unity, maximal ideals are part of a broader class of ideals known as prime ideals.

DEFINITION 3.11.12. Let R be a commutative ring. A proper ideal $\mathfrak{p}$ of R is said to be a *prime ideal* (or prime) if for all $b, c \in R$ with $bc \in \mathfrak{p}$, either $b \in \mathfrak{p}$ or $c \in \mathfrak{p}$.

EXAMPLES 3.11.13.

a. If A is an integral domain, then (0) is a prime ideal.

b. In $\mathbb{Z}$, the prime ideals are exactly (0) and the $p\mathbb{Z}$ for p prime. That is, if $ab \in (p)$ with p prime, then p divides ab, so p divides a or p divides b, and hence either $a \in (p)$ or $b \in (p)$.

We have the following analogue of Theorem 3.11.4.

THEOREM 3.11.14. *Let R be a commutative ring. Then a proper ideal $\mathfrak{p}$ of R is prime if and only if $R/\mathfrak{p}$ is an integral domain.*

PROOF. The ideal $\mathfrak{p}$ is prime if and only if $ab \in \mathfrak{p}$ implies than $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$, which translates to the fact that $ab = 0$ imples $a = 0$ or $b = 0$ in the ring $R/\mathfrak{p}$. □

COROLLARY 3.11.15. *Let R be a commutative ring. Then every maximal ideal of R is prime.*

PROOF. If $\mathfrak{m}$ is a maximal ideal of R, then Theorem 3.11.4 then tells us that $R/\mathfrak{m}$ is a field. Theorem 3.11.14 yields that $\mathfrak{m}$ is prime. □

As for polynomial rings over fields, we have the following theorem.

PROPOSITION 3.11.16. *Let F be a field. The prime ideals in $F[x]$ are exactly (0) and those (f) such that $f \in F[x]$ is irreducible.*

PROOF. Note that if f is nonconstant and reducible, then $f = gh$ for some nonconstant $g, h \in F[x]$ of degree less than $\deg f$, so $g, h \notin (f)$. Therefore, (f) is not prime.

On the other hand, if f is nonconstant and irreducible, then Proposition 3.11.3 tells us that (f) is maximal, and Corollary 3.11.15 then tells us that (f) is prime. □

EXAMPLE 3.11.17. In $\mathbb{Z}[x]$, the ideal (x) is prime, since $\mathbb{Z}[x]/(x) \cong \mathbb{Z}$, but (x) is not maximal. This follows either from the fact that $\mathbb{Z}$ is not a field, or the fact that (x) is properly contained in (p, x) for any prime p.

CHAPTER 4

Advanced group theory

4.1. Isomorphism theorems

We have already proven the first isomorphism theorem. In this section, we shall use it in deriving two others.

DEFINITION 4.1.1. Let H and K be subgroups of a group G. We define the *join* $H \vee K$ of H and K to be the subgroup of G generated by $H \cup K$.

REMARK 4.1.2. Note that the join $H \vee K$ contains (and is generated by) the set

$$HK = \{hk \mid h \in H, k \in K\}$$

and $H \vee K = HK$ and only if $HK \leqslant G$.

EXAMPLE 4.1.3. Take $H = \langle (1\ 2) \rangle$ and $K = \langle (1\ 3) \rangle$ as subgroups of S_3. We have

$$HK = \{e, (1\ 2), (1\ 3), (1\ 3\ 2)\},$$

which is not a subgroup of S_3, while $H \vee K = S_3$.

LEMMA 4.1.4. *We have that $HK \leqslant G$ if and only if $HK = KH$.*

PROOF. Suppose first that $HK \leqslant G$. Let $h \in H$ and $k \in K$. Since $h, k \in HK$, we have $kh \in HK$, as $HK \leqslant G$. Thus $KH \subseteq HK$. On the other hand, we have $(hk)^{-1} = k^{-1}h^{-1} \in KH$, so the inverse of every element of HK is contained in KH. But every element of HK is the inverse of some element of HK since $HK \leqslant G$, so $HK \subseteq KH$ as well. Thus, we have $HK = KH$.

Now suppose that $HK = KH$. We always have $e = e \cdot e \in HK$. Moreover, if $h, h' \in H$ and $k, k' \in K$, then

$$hk \cdot h'k' = h(kh')k',$$

and since $KH = HK$, there exists $h'' \in H$, $k'' \in K$ such that $kh' = h''k''$, so

$$h(kh')k' = h(h''k'')k' = hh'' \cdot k''k' \in HK.$$

Moreover, we have

$$(hk)^{-1} = k^{-1}h^{-1} \in KH,$$

but $KH = HK$, so $(hk)^{-1} \in HK$. Thus, we have that $HK \leqslant G$. □

COROLLARY 4.1.5. *Suppose that H and N are subgroups of G with N normal. Then we have $HN \leqslant G$. If H is normal in G as well, then we have $HN \trianglelefteq G$.*

PROOF. By Lemma 4.1.4, it suffices to show that $HN = NH$. But $N \trianglelefteq G$, so $hN = Nh$ for all $h \in H$, which means that

$$HN = \bigcup_{h\in H} hN = \bigcup_{h\in H} Nh = NH.$$

Moreover, if $H \trianglelefteq G$, then for any $g \in G$, we have

$$gHNg^{-1} = gHg^{-1} \cdot gNg^{-1} = HN.$$

□

THEOREM 4.1.6 (Second Isomorphism Theorem). *Let H be a subgroup of a group G, and let N be a normal subgroup of G. Then we have an isomorphism*

$$H/(H\cap N) \xrightarrow{\sim} HN/N.$$

PROOF. Define

$$\phi\colon H \to HN/N, \qquad \phi(h) = hN.$$

Then

$$\ker\phi = \{h \in H \mid h \in N\} = H\cap N.$$

Moreover, if $h \in H$ and $n \in N$, then $hnN = hN = \phi(h)$, so ϕ is surjective. The result therefore follows by the first isomorphism theorem. □

EXAMPLE 4.1.7. Consider the subgroups $H = \mathbb{Z}\times\mathbb{Z}\times\{0\}$ and $N = \{0\}\times\mathbb{Z}\times\mathbb{Z}$ of $G = \mathbb{Z}^3$. We have $HN = G$, and $H\cap N = \{0\}\times\mathbb{Z}\times\{0\}$. Note that

$$H/(H\cap N) = \frac{\mathbb{Z}\times\mathbb{Z}\times\{0\}}{\{0\}\times\mathbb{Z}\times\{0\}} \xrightarrow{\sim} \mathbb{Z}$$

via the map that takes $(a,b,0)(H\cap N)$ to a. On the other hand, we have

$$HN/N = \frac{\mathbb{Z}\times\mathbb{Z}\times\mathbb{Z}}{\{0\}\times\mathbb{Z}\times\mathbb{Z}} \xrightarrow{\sim} \mathbb{Z}$$

via the map that takes $(a,b,c)N$ to a.

REMARK 4.1.8. Suppose that H and K are subgroups of a group G with $K \leqslant H$. If K and H are both normal subgroups of G, then $K \trianglelefteq H$. On the other hand, the property of being a normal subgroup is not transitive. One may have $K \trianglelefteq H$ and $H \trianglelefteq G$ but $K \ntrianglelefteq G$!

EXAMPLE 4.1.9. Take $G = A_4$,

$$H = \langle (1\ 2)(3\ 4), (1\ 3)(2\ 4)\rangle,$$

and $K = \langle (1\ 2)(3\ 4)\rangle$. Since

$$(1\ 2)(3\ 4)\cdot(1\ 3)(2\ 4) = (1\ 4)(2\ 3) = (1\ 3)(2\ 4)\cdot(1\ 2)(3\ 4),$$

the group H is an abelian group of order 4 (isomorphic to the Klein four-group) consisting of the three cycles of cycle type $(2,2)$ and the identity. Now $K \lhd H$ since H is abelian, and $H \lhd G$ since conjugation preserves cycle type. On the other hand, K is not a normal subgroup of G since

$$(1\ 2\ 3)\cdot(1\ 2)(3\ 4)\cdot(1\ 2\ 3)^{-1} = (1\ 4)(2\ 3) \notin K.$$

THEOREM 4.1.10 (Third Isomorphism Theorem). *Let H and K be normal subgroups of a group G with $K \leqslant H$. Then we have an isomorphism*

$$G/H \xrightarrow{\sim} (G/K)/(H/K).$$

PROOF. We first remark that $H/K \leqslant G/K$ since it is a subset of G/K that is a group under the operation on G/K. Moreover, $H/K \trianglelefteq G/K$ since if $h \in H$ and $a \in G$, then $aha^{-1} = h'$ for some $h' \in H$, so

$$aK \cdot hK \cdot a^{-1}K = h'K \in H/K.$$

We may now define

$$\theta \colon G \to (G/K)/(H/K)$$

by

$$\theta(a) = (aK)(H/K).$$

By the group laws on G/K and $(G/K)/(H/K)$, we have

$$\theta(ab) = (abK)(H/K) = (aK \cdot bK)(H/K) = (aK)(H/K) \cdot (bk)(H/K) = \theta(a)\theta(b).$$

Then $\theta(a) = H/K$ if and only if $aK = hK$ for some $h \in H$, so if and only if $a \in H$. Thus $\ker\theta = H$. On the other hand, θ is surjective by definition. The result now follows from the first isomorphism theorem. □

REMARK 4.1.11. For H, K and G as in the third isomorphism theorem, the composite map

$$G \to G/K \to (G/K)/(H/K) \to G/H,$$

where the first and second maps are quotient maps and the third is the inverse of the isomorphism in the third isomorphism theorem, is exactly the quotient map $G \to G/H$.

EXAMPLE 4.1.12. Let $G = \mathbb{Z}$, $H = m\mathbb{Z}$ and $K = n\mathbb{Z}$, where $m, n \geq 1$ and m divides n, so that $K\langle H$. Then $G/H = \mathbb{Z}/m\mathbb{Z}$, $G/K = \mathbb{Z}/n\mathbb{Z}$, and $H/K = m\mathbb{Z}/n\mathbb{Z}$. We note that

$$(G/K)/(H/K) = \frac{\mathbb{Z}/n\mathbb{Z}}{m\mathbb{Z}/n\mathbb{Z}} = \frac{\mathbb{Z}/n\mathbb{Z}}{\langle m\rangle} \xrightarrow{\sim} \mathbb{Z}/m\mathbb{Z} = G/H,$$

the map in the last isomorphism being induced by the natural reduction-modulo-m map from $\mathbb{Z}/n\mathbb{Z}$ to $\mathbb{Z}/m\mathbb{Z}$ and the first isomorphism theorem.

EXAMPLE 4.1.13. Let $G = \mathbb{Z}^3$, $H = \mathbb{Z} \times \mathbb{Z} \times \{0\}$ and $K = \mathbb{Z} \times \{0\} \times \{0\}$. Then $G/H \cong \mathbb{Z}$ via the map that takes $(a,b,c)H$ to c, while $G/K \cong \mathbb{Z} \times \mathbb{Z}$ via the map that takes $(a,b,c)K$ to (b,c), and H/K has image $\mathbb{Z} \times \{0\}$ under this map. Then

$$(G/K)/(H/K) \xrightarrow{\sim} \frac{\mathbb{Z} \times \mathbb{Z}}{\mathbb{Z} \times \{0\}} \xrightarrow{\sim} \mathbb{Z},$$

where the latter map takes $(b,c)(\mathbb{Z} \times \mathbb{Z})$ to c.

We also have the following, known as the butterfly (or Zassenhaus) lemma, which we state without proof.

THEOREM 4.1.14 (Butterfly lemma). *Let H, K, A, B be subgroups of a group G with $A \trianglelefteq H$ and $B \trianglelefteq K$. Then there is a canonical isomorphism*

$$\frac{A(H \cap K)}{(A(H \cap B)} \cong \frac{B(H \cap K)}{B(A \cap K)}.$$

4.2. Commutators and simple groups

DEFINITION 4.2.1. Let G be a group and $a, b \in G$. The *commutator* of a and b is

$$[a,b] = aba^{-1}b^{-1}.$$

DEFINITION 4.2.2. The *commutator subgroup* $[G,G]$ of a group G is the subgroup of G generated by its commutators, which is to say

$$[G,G] = \langle [a,b] \mid a,b \in G \rangle.$$

REMARK 4.2.3. If G is an abelian group, then $[G,G] = \{e\}$.

EXAMPLE 4.2.4. In D_n, we have

$$[r^i, r^j s] = r^i(r^j s)(r^{-i})(sr^{-j}) = r^{2i},$$
$$[r^i s, r^j s] = (r^i s)(r^j s)(sr^{-i})(sr^{-j}) = r^{i-j} s r^{j-i} s = r^{2(i-j)}.$$

Therefore, we have that

$$[D_n, D_n] = \langle r^2 \rangle,$$

which has index 2 and 4 in D_n in the cases that n is odd and even, respectively.

EXAMPLE 4.2.5. We have

$$[\mathrm{GL}_n(\mathbb{R}), \mathrm{GL}_n(\mathbb{R})] \subseteq \mathrm{SL}_n(\mathbb{R}).$$

since $\det(ABA^{-1}B^{-1}) = 1$ for any $A, B \in \mathrm{GL}_n(\mathbb{R})$. The opposite equality also holds, but we shall not prove it here.

LEMMA 4.2.6. *The commutator subgroup of G is a normal subgroup of G.*

PROOF. Let $a, b, g \in G$. We have

$$g[a,b]g^{-1} = gaba^{-1}b^{-1}g^{-1} = (ga)b(ga)^{-1}b^{-1} \cdot bgb^{-1}g^{-1} = [ga,b][b,g] \in [G,G].$$

Since every element of $[G,G]$ is a product of powers of elements of the form $[a,b]$ with $a, b \in G$ and every such element is sent to another element of $[G,G]$ by the conjugation homomorphism γ_g, the image of γ_g is contained in $[G,G]$. Since this holds for all $g \in G$, we have that $[G,G]$ is normal. □

EXAMPLE 4.2.7. Let $n \geq 3$. We claim that $[S_n, S_n] = A_n$. Note that

$$[\sigma, (a\ b)] = \sigma(a\ b)\sigma^{-1}(a\ b) = (\sigma(a)\ \sigma(b))(a\ b)$$

for $\sigma \in S_n$ and $a, b \in X_n$ with $a < b$. It follows that $[G,G]$ contains all products of two transposiitons in S_n. Moreover, these generate A_n by definition, so $A_n \subseteq [S_n, S_n]$. Furthermore, every element of $[S_n, S_n]$ is even as every such element is a product of elements of the form $[\sigma, \tau]$ with $\sigma, \tau \in S_n$, and these satisfy

$$\operatorname{sign}([\sigma, \tau]) = \operatorname{sign}(\sigma)\operatorname{sign}(\tau)\operatorname{sign}(\sigma)^{-1}\operatorname{sign}(\tau)^{-1} = 1.$$

THEOREM 4.2.8. *Let N be a normal subgroup of G. Then G/N is an abelian group if and only if $[G,G] \leqslant N$.*

PROOF. Let $a, b \in G$. We have $abN = baN$ if and only if $a^{-1}b^{-1}ab \in N$, so if and only if $[a^{-1}, b^{-1}] \in N$. But $[G,G]$ is the smallest subgroup of G containing $[c,d]$ for every $c, d \in G$, so N is normal if and only if $[G,G]$ is contained in N. □

DEFINITION 4.2.9. The *maximal abelian quotient*, or *abelianization*, G^{ab} of a group G is the quotient group

$$G^{\mathrm{ab}} = G/[G,G].$$

We have the following consequence of Theorem 4.2.8.

COROLLARY 4.2.10. *Let G be a group and H be an abelian group, and suppose $\phi \colon G \to H$ is a homomorphism. Then there exists a homomorphism $\bar{\phi} \colon G^{\mathrm{ab}} \to H$ with $\bar{\phi} \circ \pi_{[G,G]} = \phi$, where $\pi_{[G,G]} \colon G \to G^{\mathrm{ab}}$ is the quotient map.*

PROOF. By the first isomorphism theorem, there exists a unique map $\psi \colon G/\ker\phi \to H$ with $\psi \circ \pi_{\ker\phi} = \phi$. By Theorem 4.2.8, we have that $[G,G] \subseteq \ker\phi$. and now the third isomorphism theorem provides a composite map

$$\beta \colon G^{\mathrm{ab}} = G/[G,G] \to (G/[G,G])/(\ker\phi/[G,G]) \xrightarrow{\sim} G/\ker\phi$$

such that $\pi_{\ker\phi} = \beta \circ \pi_{[G,G]}$. Set $\bar{\phi} = \psi \circ \beta$. Then

$$\bar{\phi} \circ \pi_{[G,G]} = \psi \circ \beta \circ \pi_{[G,G]} = \psi \circ \pi_{\ker\phi} = \phi,$$

as desired. □

EXAMPLE 4.2.11. The abelianization of D_n is $D_n/\langle r^2 \rangle$, and if n is even this group is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ under the map $\bar{\phi}$ induced by the homomorphism

$$\phi \colon D_n \to \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$$

that takes $r^i s^j$ to (i, j). If n is odd, then $D_n/\langle r^2 \rangle$ is isomorphic to $\mathbb{Z}/2\mathbb{Z}$ via the map from D_n that takes $r^i s^j$ to j.

EXAMPLE 4.2.12. We have $S_n^{\mathrm{ab}} = S_n/A_n$, and so $S_n^{\mathrm{ab}} \cong \mathbb{Z}/2\mathbb{Z}$ via the map induced by the sign map.

Here is another nice class of normal subgroups.

DEFINITION 4.2.13. The *center* of a group G is the subgroup

$$Z(G) = \{z \in G \mid za = az \text{ for all } a \in G\}.$$

Since $az = za$ for all $z \in Z(G)$ and $a \in G$, we clearly have that $aZ(G) = Z(G)a$ for all $a \in G$, and hence we have that $Z(G) \trianglelefteqslant G$. We leave the verification of the following examples to the reader.

EXAMPLES 4.2.14.

a. If G is abelian, then $Z(G) = G$.

b. For $n \geq 3$, we have $Z(S_n) = \{e\}$.

c. For $n \geq 3$, we have $Z(D_n) = \langle r^{n/2} \rangle$ if n is even and $Z(D_n) = \{e\}$ if n is odd.

d. For $n \geq 2$, we have that $Z(\mathrm{GL}_n(\mathbb{R}))$ is the subgroup of scalar matrices.

DEFINITION 4.2.15. A nontrivial group G is called *simple* if it has no nontrivial, improper normal subgroups.

EXAMPLE 4.2.16. An abelian group G is simple if and only if it is cyclic of prime order, since otherwise it will have a nontrivial, improper subgroup, which is automatically normal since G is abelian.

EXAMPLES 4.2.17. The groups S_n and D_n for $n \geq 3$ are not simple, since they contain improper, nontrivial normal subgroups. Moreover, A_4 is not simple, as it contains the normal subgroup $\langle (1\ 2)(3\ 4), (1\ 3)(2\ 4) \rangle$ of order 8, as seen in Example 4.1.9.

We have the following easy lemma.

LEMMA 4.2.18. *If G is simple and nonabelian, then $Z(G) = \{e\}$.*

PROOF. If G is nonabelian, then $Z(G) \neq G$, and if G is also simple, then since $Z(G)$ is normal, we must have $Z(G) = \{e\}$. □

We note that if a group is not simple, we can find a nontrivial quotient of it by a nontrivial normal subgroup that is.

DEFINITION 4.2.19. We say that a normal subgroup M of a group G is a *maximal normal subgroup* if it is not contained in any larger proper normal subgroup of G.

EXAMPLE 4.2.20. Any subgroup of index 2 in a group is a maximal normal subgroup, since such a subgroup is normal and is not contained in a larger proper normal subgroup, being that its index would have to be smaller than 2, but greater than 1.

PROPOSITION 4.2.21. *Let N be a normal subgroup of G. Then G/N is simple if and only if N is maximal.*

PROOF. This is an immediate consequence of Proposition 2.13.10, since G/N is simple if and only if it has no proper normal subgroups, which are in bijection with the proper normal subgroups of G containing N. □

EXAMPLE 4.2.22. Since A_n is a maximal normal subgroup of S_n, the quotient S_n/A_n is simple (which we already knew since it is isomorphic to $\mathbb{Z}/2\mathbb{Z}$).

Finite simple groups are considered the building blocks of all finite groups. Their classification was the major project in group theory during the 20th century, and it was finally finished at the beginning of the 21st. Many examples of nonabelian finite simple groups are simple enough to give, though proving they are simple is another matter.

EXAMPLE 4.2.23. The groups A_n are simple for all $n \geq 5$. In fact, A_5 is a finite simple group with the smallest possible order, which is 60, and it is the unique such group up to isomorphism. We defer the proofs of these facts until later.

We mention one more broad class of examples of finite simple groups.

EXAMPLE 4.2.24. We remark that $\mathbb{Z}/p\mathbb{Z}$ has two binary operations of addition and multiplication, and these satsify the distributive property. Hence we may consider the set $\mathrm{GL}_n(\mathbb{Z}/p\mathbb{Z})$ of invertible n by n matrices with entries in $\mathbb{Z}/p\mathbb{Z}$, and this forms a group under multiplication. We also have its subgroup $\mathrm{SL}_n(\mathbb{Z}/p\mathbb{Z})$ of matrices with determinant 1. It is not necessarily simple, as it is possible that it can have nontrivial center: the group of scalar matrices with determinant 1. I.e., $aI \in \mathrm{SL}_n(\mathbb{Z}/p\mathbb{Z})$ if and only if $a^n = 1$ in $\mathbb{Z}/p\mathbb{Z}$. The quotient of $\mathrm{SL}_n(\mathbb{Z}/p\mathbb{Z})$ by its center is called $\mathrm{PSL}_n(\mathbb{Z}/p\mathbb{Z})$. It turns out that $\mathrm{PSL}_n(\mathbb{Z}/p\mathbb{Z})$ is simple for all primes p for all $n \geq 3$ and for all primes $p \geq 5$ and $n = 2$.

4.3. Automorphism groups

DEFINITION 4.3.1. An *automorphism* of a group G is an isomorphism $\phi\colon G \to G$.

The subgroup test shows quickly that the set of automorphisms of G forms a subgroup of G under composition.

PROPOSITION 4.3.2. *The set of automorphisms* $\mathrm{Aut}(G)$ *of a group G forms a group under composition.*

PROOF. Since composition of functions is associative, to check that $\mathrm{Aut}(G)$ is a group, we need only check that it contains an identity element, which it clearly does, and that it contains inverses, which is does since the inverse of an isomorphism is an isomorphism. □

DEFINITION 4.3.3. The *automorphism group* $\mathrm{Aut}(G)$ of a group G is the group of automorphisms of G under conjugation.

EXAMPLES 4.3.4.

a. We have $\mathrm{Aut}(\mathbb{Z}/n\mathbb{Z}) \cong (\mathbb{Z}/n\mathbb{Z})^\times$ for $n \geq 1$ via the map that takes $\phi \in \mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$ to $\phi(1)$. In fact, we have $\phi(a) = a\phi(1)$ for any $a \in \mathbb{Z}/n\mathbb{Z}$, and so ϕ is multiplication by $\phi(1)$. This can be invertible if and only if a is a unit in $\mathbb{Z}/n\mathbb{Z}$.

b. The same discussion as in part a tells us that $\mathrm{Aut}(\mathbb{Z}) = \mathbb{Z}^\times = \langle -1 \rangle$.

c. We have $\mathrm{Aut}(\mathbb{Z}^n) \cong \mathrm{GL}_n(\mathbb{Z})$. That is, if $\phi \in \mathrm{Aut}(\mathbb{Z}^n)$ and e_i is the ith element in the standard basis of $\mathbb{Z}^n$, then $\phi(e_i)$ determines the ith c olumn of a matrix in $\mathrm{GL}_n(\mathbb{Z})$. The inverse map is given by allowing $\mathrm{GL}_n(\mathbb{Z})$ to act on $\mathbb{Z}^n$ by left multiplication, viewing an element of $\mathbb{Z}^n$ as a column vector.

We give the example of the automorphisms of the dihedral group as a proposition.

PROPOSITION 4.3.5. *For $n \geq 3$, the group* $\mathrm{Aut}(D_n)$ *is isomorphic to the subgroup* $\mathrm{Aff}(\mathbb{Z}/n\mathbb{Z})$ *of* $\mathrm{GL}_2(\mathbb{Z}/n\mathbb{Z})$ *given by*

$$\left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \middle| a \in (\mathbb{Z}/n\mathbb{Z})^\times, b \in \mathbb{Z}/n\mathbb{Z} \right\}.$$

PROOF. The isomorphism $f\colon \operatorname{Aff}(\mathbb{Z}/n\mathbb{Z}) \to \operatorname{Aut}(D_n)$ is given by

$$f\colon \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \mapsto \phi_{a,b},$$

where $\phi_{a,b}(r) = r^a$ and $\phi_{a,b}(s) = r^b s$. Since D_n is generated by r and s, there exists at most one element of $\operatorname{Aut}(D_n)$ taking these values on r and s. Since $F_{r,s}$ is free, we can define $\Phi\colon F_{r,s} \to D_n$ by $\Phi(r) = r^a$ and $\Phi(s) = r^b s$ for a and b as above. Note that $\Phi(r^n) = r^{an} = e$, $\Phi(s^2) = (r^b s)^2 = e$, and $\Phi((rs)^2) = (r^{a+b}s)^2 = e$, so by the presentation $D_n \cong \langle r, s \mid r^n, s^2, rsrs\rangle$, we have the existence of $\phi_{a,b}$. In that a is invertible modulo n, we have

$$\langle r^a, r^b s\rangle = \langle r, r^b s\rangle = \langle r, s\rangle = D_n,$$

so $\phi_{a,b}$ is onto and hence in $\operatorname{Aut}(D_n)$ as D_n is finite.

Now, any $\phi \in \operatorname{Aut}(D_n)$ must send r to another element of order n, so r^a with $a \in \mathbb{Z}$ prime to n. It must also send s to an element of order 2 that cannot be in the subgroup $\langle \phi(r)\rangle = \langle r\rangle$, since ϕ is surjective. Thus, $\phi(s) = r^b s$ for some $b \in \mathbb{Z}$. Thus, f is onto, and it is one-to-one by definition. To see it is a homomorphism, note that

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a' & b' \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} aa' & b+ab' \\ 0 & 1 \end{pmatrix},$$

while

$$\phi_{a,b}(\phi_{a',b'}(r)) = \phi_{a,b}(r^{a'}) = r^{aa'} \quad \text{and} \quad \phi_{a,b}(\phi_{a',b'}(s)) = \phi_{a,b}(r^{b'}s) = r^{ab'+b}s.$$

□

DEFINITION 4.3.6. An automorphism of G is called an *inner automorphism* (or *inner*) if it is equal to a conjugation map $\gamma_a\colon G \to G$ for some $a \in G$.

LEMMA 4.3.7. *The set* $\operatorname{Inn}(G)$ *of inner automorphisms of G is a subgroup of* $\operatorname{Aut}(G)$ *under composition.*

PROOF. That the inner automorphisms form a subgroup amounts to the facts that $\gamma_e = \mathrm{id}_G$, that $\gamma_{ab} = \gamma_a\gamma_b$, and that $\gamma_{a^{-1}} = \gamma_a^{-1}$ for $a \in G$. □

DEFINITION 4.3.8. The *inner automorphism group* $\operatorname{Inn}(G)$ is the subgroup of G consisting of inner automorphisms.

LEMMA 4.3.9. *For a group G, the inner automorphism group* $\operatorname{Inn}(G)$ *is a normal subgroup of G.*

PROOF. For $\phi \in \operatorname{Aut}(G)$ and $g, x \in G$, we have

$$(\phi \circ \gamma_g \circ \phi^{-1})(x) = \phi(g\phi^{-1}(x)g^{-1}) = \phi(g)x\phi(g)^{-1} = \gamma_{\phi(g)}(x),$$

so $\phi\gamma_g\phi^{-1} = \gamma_{\phi(g)}$ lies in $\operatorname{Inn}(G)$. □

DEFINITION 4.3.10. The *outer automorphism group* of G is the quotient group $\operatorname{Out}(G) = \operatorname{Aut}(G)/\operatorname{Inn}(G)$.

REMARK 4.3.11. An automorphism is sometimes called *outer* if it is not inner. However, the outer automorphism group is not a group of automorphisms, but rather cosets thereof.

REMARK 4.3.12. If G is an abelian group, then every inner automorphism of G is trivial, so $\mathrm{Out}(G) \cong \mathrm{Aut}(G)$.

EXAMPLE 4.3.13. The group $\mathrm{Inn}(D_n)$ for $n \geq 3$ is generated by the images γ_r and γ_s of r and s under $\gamma\colon D_n \to \mathrm{Aut}(D_n)$. We have $\gamma_r(s) = r^2 s$ and $\gamma_s(r) = r^{-1}$, and of course $\gamma_r(r) = r$ and $\gamma_s(s) = s$. Using the isomorphism of Proposition 4.3.5, we that $\mathrm{Inn}(G)$ is isomorphic to the subgroup of $\mathrm{GL}_2(\mathbb{Z}/n\mathbb{Z})$ given by

$$K = \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \middle| a = \pm 1, b \in 2\mathbb{Z}/n\mathbb{Z} \right\}.$$

The quotient group $\mathrm{Out}(D_n)$ is then in bijection with pairs $(i, j) \in (\mathbb{Z}/n\mathbb{Z})^\times/\langle -1 \rangle \times \mathbb{Z}/(n,2)\mathbb{Z}$. We leave it to the reader to check that

$$\mathrm{Out}(D_n) \cong \begin{cases} (\mathbb{Z}/n\mathbb{Z})^\times/\langle -1 \rangle & \text{if } n \text{ is odd,} \\ (\mathbb{Z}/n\mathbb{Z})^\times/\langle -1 \rangle \times \mathbb{Z}/2\mathbb{Z} & \text{if } n \text{ is even} \end{cases}$$

using the fact that $\left(\begin{smallmatrix} 1 & n/2 \\ 0 & 1 \end{smallmatrix}\right)$ is in the center of the group H of Proposition 4.3.5 if n is even.

DEFINITION 4.3.14. A subgroup N of a group G is *characteristic* if $\phi(N) = N$ for every $\phi \in \mathrm{Aut}(G)$.

LEMMA 4.3.15. *Let G be a group.*

a. If H is the unique subgroup of G of a given order, then H is characteristic.

b. The center $Z(G)$ of a group G is characteristic.

c. The commutator subgroup $[G, G]$ of a group G is characteristic.

PROOF. Let $\phi \in \mathrm{Aut}(G)$. For part a, note that $\phi(H)$ has the same order as H. For part b, note that $\phi(a)\phi(x) = \phi(x)\phi(a)$ for any $a \in Z(G)$ and $x \in G$, but ϕ is onto, so $\phi(a)$ commutes with every element of G. So, $\phi\colon Z(G) \to Z(G)$, and $\phi^{-1} \in \mathrm{Aut}(G)$ has the same property, so $\phi(Z(G)) = Z(G)$. For part c, note that $[G, G]$ is generated by commutators $[a, b]$ with $a, b \in G$, and $\phi([a, b]) = [\phi(a), \phi(b)] \in [G, G]$. We can see that $\phi([G, G])$ actually equals $[G, G]$ by noting that ϕ is onto. □

LEMMA 4.3.16. *If K is a characteristic subgroup of a normal subgroup N of a group G, then $K \trianglelefteq G$. If, moreover, N is characteristic in G, then K is characteristic in G.*

PROOF. Let $a \in G$. Then the restriction of γ_a to N provides an element of $\mathrm{Aut}(N)$ as N is normal, and so $aKa^{-1} = \gamma_a(K) = K$ as K is characteristic. Thus $K \trianglelefteq G$.

If N is characteristic in G and $\phi \in \mathrm{Aut}(G)$, then the restriction of ϕ to N is an automorphism of N as N is characteristic in G, and so $\phi(K) = K$ as K is characteristic in N. Thus K is characteristic in G. □

4.4. Free abelian groups

The theory of free abelian groups is the analogue of the theory of vector spaces when the scalars are taken to be not real or complex numbers, but rather integers. In this section, we briefly explore this theory.

DEFINITION 4.4.1. An abelian group G (under addition) is said to be a *free abelian group* if it has a generating set X of G such that for any $n \geq 1$, distinct $x_1, x_2, \dots, x_n \in X$, and $c_1, c_2, \dots, c_n \in \mathbb{Z}$ with

$$\sum_{i=1}^{n} c_i x_i = 0,$$

one has $c_1 = c_2 = \cdots = c_n = 0$. Such a set X is called a *basis* of G, and G is said to be free on X.

EXAMPLE 4.4.2. The group $\mathbb{Z}^n$ is free on the set $\{e_1, e_2, \dots, e_n\}$, where $e_i \in \mathbb{Z}^n$ is the tuple that is 0 in every coordinate but the ith, where it is 1.

EXAMPLE 4.4.3. The group $\mathbb{Z}/n\mathbb{Z}$ is not free for $n \geq 1$, since one has $na = 0$ for every $a \in \mathbb{Z}/n\mathbb{Z}$.

REMARK 4.4.4. Much as in linear algebra, freeness of an abelian group G on a set X implies that there is a unique way to represent any nonzero element $a \in G$ as a sum

$$a = \sum_{i=1}^{n} c_i x_i$$

for some $n \geq 1$, distinct elements $x_1, x_2, \dots, x_n$ of X, and nonzero elements $c_1, c_2, \dots, c_n$ of $\mathbb{Z}$.

DEFINITION 4.4.5. If $x_1, x_2, \dots, x_n \in G$, where G is a free abelian group, then we refer to a sum

$$\sum_{i=1}^{n} c_i x_i$$

with $c_1, c_2, \dots, c_n \in \mathbb{Z}$ as an *integral linear combination* of elements of G.

Let us begin with a very general construction of a direct sum of groups, which we will then specialize immediately to the case of interest that the groups are all $\mathbb{Z}$.

DEFINITION 4.4.6. Let I be an indexing set and $\{G_i \mid i \in I\}$ a collection of abelian groups. Let

$$\bigoplus_{i \in I} G_i = \Big\{ (a_i)_{i \in I} \in \prod_{i \in I} G_i \mid a_i \in G_i,\, a_i = 0 \text{ for all but finitely many } i \in I \Big\}.$$

Then $\bigoplus_{i \in I} G_i$ is a subgroup of $\prod_{i \in I} G_i$ known as the *direct sum* of the groups G_i.

REMARK 4.4.7. When I is finite, we have $\bigoplus_{i \in I} G_i = \prod_{i \in I} G_i$.

NOTATION 4.4.8. The symbol $\delta_{i,j}$ (or δ_{ij}), for i and j in some set I, is taken to mean

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

EXAMPLE 4.4.9. For any indexing set I, the direct sum

$$\bigoplus_{i\in I}\mathbb{Z} = \Big\{(a_i)_{i\in I} \in \prod_{i\in I}\mathbb{Z} \mid a_i \in \mathbb{Z},\, a_i = 0 \text{ for all but finitely many } i \in I\Big\}.$$

is a free group under coordinate-wise addition with basis $\{e_j = (\delta_{ij})_{i\in I} \mid j \in I\}$.

DEFINITION 4.4.10. The basis $\{e_i \mid i \in I\}$ in Example 4.4.9 is known as the *standard basis* of $\bigoplus_{i\in I}\mathbb{Z}$.

Free groups have the property that homomorphisms are defined uniquely by their values on a basis, as we now show.

PROPOSITION 4.4.11. *Let G be an abelian group. Then G is free on a subset X if and only if, for every function $\bar{\phi}\colon X \to G'$, where G' is an abelian group, there exists a unique homomorphism $\phi\colon G \to G'$ with $\phi(x) = \bar{\phi}(x)$ for all $x \in X$.*

PROOF. Suppose first that G is free on a basis X. Then for $c_1, c_2, \dots, c_n \in \mathbb{Z}$ and distinct elements $x_1, x_2, \dots, x_n \in X$, define

$$\phi\Big(\sum_{i=1}^n c_i x_i\Big) = \sum_{i=1}^n c_i \bar{\phi}(x_i).$$

The map ϕ is then a well-defined map on all of G by Remark 4.4.4, and it is easy to check that it is a homomorphism. Moreover, if $\psi\colon G \to G'$ is any homomorphism with $\psi(x) = \bar{\phi}(x)$ for all $x \in X$, then

$$\psi\Big(\sum_{i=1}^n c_i x_i\Big) = \sum_{i=1}^n c_i \psi(x_i) = \sum_{i=1}^n c_i \bar{\phi}(x_i) = \phi\Big(\sum_{i=1}^n c_i x_i\Big).$$

Conversely, suppose that G and X have the property of the proposition. We claim that G is free on X. First, suppose that $x_i \in X$ and $c_i \in \mathbb{Z}$ for $1 \le i \le n$ and some $n \ge 1$ are such that

$$a = \sum_{i=1}^n c_i x_i = 0.$$

Define $\bar{\phi}\colon X \to \mathbb{Z}^n$ by $\bar{\phi}(x_i) = e_i$. Then

$$0 = \bar{\phi}(a) = \sum_{i=1}^n c_i e_i,$$

which forces $c_i = 0$ for all i, as the e_i form a basis of $\mathbb{Z}^n$.

Next, let H be the subgroup of G generated by X. We define two homomorphisms $G \to G$. One is given by id_G, while is the composition of the map $\pi\colon G \to H$ uniquely determined by $\pi(x) = x$ for all $x \in X$ with the inclusion map $\iota\colon H \to G$. By assumption, then, we must have $\mathrm{id}_G = \iota \circ \pi$, and as the latter map has image H, we have $G = H$. Thus, G is free on X. □

REMARK 4.4.12. The existence of unique homomorphisms of a free abelian group G with prescibed values on a basis X, as found in Proposition 4.4.11, is often referred to as the universal property of G.

COROLLARY 4.4.13. *Suppose that G is a free abelian group on a basis $X = \{x_i \mid i \in I\}$, where I is an indexing set. Then there is a unique isomorphism*

$$\psi\colon \bigoplus_{i\in I} \mathbb{Z} \xrightarrow{\sim} G$$

such that $\psi(e_i) = x_i$ for all $i \in I$, where $\{e_i \mid i \in I\}$ is the standard basis of G.

PROOF. We can define ψ as in the statement of the corollary by Proposition 4.4.11, which also implies the existence of a unique homomorphism $\phi\colon G \to \bigoplus_{i\in I} \mathbb{Z}$ such that $\phi(x_i) = e_i$ for all $i \in I$. Since $\phi \circ \psi(e_i) = e_i$ and $\psi \circ \phi(x_i) = x_i$ for all $i \in I$, the same proposition implies that $\phi \circ \psi$ and $\psi \circ \phi$ are the identity homomorphisms. In particular, ψ is an isomorphism. □

DEFINITION 4.4.14. The general linear group $\mathrm{GL}_n(\mathbb{Z})$ of degree n is the group of n-by-n matrices with integer entries which have inverses with integer entries, with respect to the operation of matrix multiplication.

REMARK 4.4.15. A n-by-n matrix A with integer entries has an inverse with integer entries if and only if $\det(A) = \pm 1$.

For two free abelian groups to be isomorphic, their bases must have the same cardinality. Equivalently, an abelian group cannot have bases of two different cardinalities. We prove this only in the special case of finitely generated abelian groups.

THEOREM 4.4.16. *Suppose that G is a free abelian group with basis X having n elements. Then every basis of G has n elements.*

PROOF. By Corollary 4.4.13, we have that $G \cong \prod_{i=1}^{n} \mathbb{Z}$. We then have that

$$G/2G \cong \prod_{i=1}^{n} \mathbb{Z}/2\mathbb{Z},$$

and so has order 2^n. If G had a different basis with a finite number of elements m, then $G/2G$ would have order 2^m, forcing $m = n$. On the other hand, if G had an infinite basis indexed by a set I, then the same argument would tell us that

$$G/2G \cong \bigoplus_{i\in I} \mathbb{Z}/2\mathbb{Z},$$

which is infinite, so impossible. □

DEFINITION 4.4.17. If G is a finitely generated, free abelian group, we refer to the number of elements in any basis of it as its *rank*.

We have the following analogue of the change-of-basis theorem in linear algebra.

PROPOSITION 4.4.18. *Let G be a free abelian group with basis $X = \{x_1, x_2, \ldots, x_n\}$. Then $X' = \{x_1', x_2', \ldots, x_n'\}$ is also a basis of G if and only if there exists a matrix $A = (a_{ij}) \in \mathrm{GL}_n(\mathbb{Z})$ such that*

$$x_i' = \sum_{j=1}^{n} a_{ij} x_j$$

for each $1 \le i \le n$.

PROOF. Since X generates G, we may write each x_i' as

$$x_i' = \sum_{j=1}^{n} a_{ij} x_j$$

for some $a_{ij} \in \mathbb{Z}$ and then form an n-by-n matrix $A = (a_{ij})$. If X' also generates G, then we may write

$$x_i = \sum_{j=1}^{n} b_{ij} x_j' = \sum_{j=1}^{n} \sum_{k=1}^{n} b_{ij} a_{jk} x_k$$

for some $b_{ij} \in \mathbb{Z}$ and then form $B = (b_{ij})$. Since X is a basis, this tells us that $BA = I$, so $A \in \mathrm{GL}_n(\mathbb{Z})$.

Conversely, if there exists a B with $BA = I$, then

$$x_i = \sum_{k=1}^{n} \Big(\sum_{j=1}^{n} b_{ij} a_{jk} \Big) x_k = \sum_{j=1}^{n} b_{ij} x_j',$$

so the x_j' generate G, and moreover they form a basis as, if

$$\sum_{i=1}^{n} c_i x_i' = 0,$$

we then have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} c_i x_j = 0,$$

so

$$\sum_{i=1}^{n} a_{ij} c_i = 0$$

for each i, or in other words the vector $c = (c_1, c_2, \dots, c_n)$ satisfies $Ac = 0$, which means $BAc = 0$, or $c = 0$. □

4.5. Finitely generated abelian groups

We begin with the following corollary of Theorem 4.4.18.

LEMMA 4.5.1. *Suppose that $X = \{x_1, x_2, \dots, x_n\}$ is a basis of a free abelian group G, and let $c_i \in \mathbb{Z}$ for $2 \leq i \leq n$. Then $X' = \{x_1', x_2, \dots, x_n\}$ with*

$$x_1' = x_1 + c_2 x_2 + \cdots + c_n x_n$$

is also a basis of G.

PROOF. Take $A \in \mathrm{GL}_n(\mathbb{Z})$ to be $A = I + \sum_{k=2}^{n} c_j E_{1j}$, where E_{ij} is the n-by-n matrix with exactly one nonzero entry, which is a 1 in the ith row and jth column. It is easy to see that $\det(A) = 1$, so $A \in \mathrm{GL}_n(\mathbb{Z})$. We then apply Proposition 4.4.18. □

We are now ready to prove the following result.

LEMMA 4.5.2. *Let G be a finitely generated, free abelian group of rank n, and let H be a nontrivial subgroup. Then there is an isomorphism*

$$\varphi\colon G \to \mathbb{Z} \times G',$$

where G' is a subgroup of G that is free abelian of rank $n-1$, such that

$$\varphi(H) = d\mathbb{Z} \times H',$$

for some $d \geq 1$, where $H' = H \cap G'$.

PROOF. Consider the set $\mathscr{B}$ of all bases of G. Let $d \geq 1$ be minimal such that there exists $X' = \{x', x_2, \dots, x_n\} \in \mathscr{B}$ such that

$$y = dx' + \sum_{i=2}^{n} d_j x_j \in H$$

for some $d_2, \dots, d_n \in \mathbb{Z}$, and fix such an X' and y. We may divide each d_i for $2 \leq i \leq n$ by d to obtain

$$d_i = q_i d + r_i$$

with $q_i \in \mathbb{Z}$ and $0 \leq r_i < n$. Then

$$y = d(x' + q_2x_2 + \cdots + q_nx_n) + r_2x_2 + \cdots + r_nx_n.$$

Let $x_1 = x' + q_2x_2 + \cdots + q_nx_n$. Then $X = \{x_1, x_2, \dots, x_n\}$ is a basis by Lemma 4.5.1. The minimality of d now forces $r_2 = \cdots = r_n = 0$. In other words, we have $y = dx_1 \in H$.

Let $G' = \langle x_2, \dots, x_n \rangle$, which is free abelian of rank $n-1$, and define a homomorphism

$$\varphi\colon G \to \mathbb{Z} \times G', \qquad \varphi(a) = (c_1, a - c_1x_1),$$

for $a \in G$, where $c_1 \in \mathbb{Z}$ is such that

$$a = \sum_{i=1}^{n} c_i x_i$$

for some $c_2, \dots, c_n \in \mathbb{Z}$. We have that $\varphi(a) = 0$ if and only if $c_1 = 0$ and $a - c_1x_1 = 0$ by definition, which occurs exactly when $a = 0$ as X is a basis. Therefore, φ is injective. Moreover, for $a' \in G'$ and $c \in \mathbb{Z}$, we clearly have that

$$\varphi(cx_1 + a') = (c, a'),$$

so φ is surjective. Therefore, φ is an isomorphism.

Finally, we compute $\varphi(H)$. Suppose $a \in H$ is written as above, and let $c_1 = qd + r$ with $q \in \mathbb{Z}$ and $0 \leq r < d$. Then $b = a - qdx_1 \in H$, and

$$b = rx_1 + \sum_{i=2}^{n} c_i x_i.$$

By the minimality of d, we must have $r = 0$. In other words, we have $b \in H'$ and

$$a = q(dx_1) + b,$$

so $\varphi(a) \in d\mathbb{Z} \times H'$. Conversely, if $(m, b) \in d\mathbb{Z} \times H'$, then

$$(m, b) = \varphi(mx_1 + b),$$

and $mx_1 \in H$ since d divides m, so $mx_1 + b \in H$. Therefore, $\varphi(H) = d\mathbb{Z} \times H'$, as desired. □

We also note the following easy corollary of Theorem 2.5.17, obtained by applying it recursively.

COROLLARY 4.5.3. *Let m be a positive integer, and for some $k \geq 0$, write*

$$m = p_1^{r_1} p_2^{r_2} \cdots p_k^{r_k}$$

for distinct prime numbers $p_1, p_2, \dots, p_k$ and $r_1, r_2, \dots, r_k \geq 2$. Then

$$\mathbb{Z}/m\mathbb{Z} \cong \mathbb{Z}/p_1^{r_1}\mathbb{Z} \times \mathbb{Z}/p_2^{r_2}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_k^{r_k}\mathbb{Z}.$$

We can now classify the finitely generated abelian groups up to isomorphism.

THEOREM 4.5.4 (Structure theorem for finitely generated abelian groups). *Let G be a finitely generated abelian group. Then there exist $k, r \geq 0$ and positive integers $d_1, d_2, \dots, d_k \geq 2$ such that there is an isomorphism*

$$G \cong \mathbb{Z}^r \times (\mathbb{Z}/d_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/d_k\mathbb{Z}).$$

In fact, the integers d_i may be chosen so that d_{i+1} divides d_i for each $1 \leq i \leq k-1$, and then these are the unique r, k, and $d_1, d_2, \dots, d_k$ with those properties. Alternatively, we may choose the isomorphism so that each d_i is a power of a prime number, in which case the decomposition is again unique up to reordering.

PROOF. We prove the result by induction on the number of elements n in a finite generating set of G, where we may consider the trivial group to be generated by the empty set. The case $n = 0$ is then just the case that $r = k = 0$, and we have the result. Suppose we know the result for all abelian groups that can be generated by n elements. Let G be an abelian group for which $X = \{x_1, x_2, \dots, x_{n+1}\}$ is a minimal set of generators. Then there exists a unique surjective homomorphism $\psi \colon \mathbb{Z}^{n+1} \to G$ such that $\psi(e_i) = x_i$ for $1 \leq i \leq n+1$. Let $H = \ker \psi \leqslant \mathbb{Z}^{n+1}$.

By Lemma 4.5.2, we have an isomorphism

$$\varphi \colon \mathbb{Z}^{n+1} \to \mathbb{Z} \times \mathbb{Z}^n$$

such that $\varphi(H) = d\mathbb{Z} \times H'$ for some $H' \leqslant \mathbb{Z}^n$ and $d \geq 1$. By the first isomorphism theorem, we have

$$G \cong \mathbb{Z}^{n+1}/H \cong \frac{\mathbb{Z} \times \mathbb{Z}^n}{d\mathbb{Z} \times H'} \cong \mathbb{Z}/d\mathbb{Z} \times \mathbb{Z}^n/H'.$$

Now, since $\mathbb{Z}^n/H'$ can be generated by n elements, it may be written by induction as

$$\mathbb{Z}^n/H' \cong \mathbb{Z}^r \times (\mathbb{Z}/d_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/d_{k-1}\mathbb{Z})$$

for some $d_1, d_2, \dots, d_{k-1} \geq 2$ with $r + k - 1 \leq n$. Letting $d_k = d$ if $d \geq 2$ and noting that $G \cong \mathbb{Z}^n/H'$ if $d = 1$, we therefore have the first statement of the theorem.

By Corollary 4.5.3, we may can decompose each $\mathbb{Z}/d_i\mathbb{Z}$ into a finite direct product of groups of the form $\mathbb{Z}/p^k\mathbb{Z}$ with p prime and $k \geq 2$, proving the last decomposition. On the other hand, suppose we have decomposed G up to isomorphism as

$$G \cong \mathbb{Z}^r \times P_1 \times P_2 \cdots \times P_t,$$

where $p_1, p_2, \dots, p_t$ are distinct prime numbers and each P_i for $1 \le i \le t$ is a finite abelian p_i-group, which in turn we have written as

$$P_i \cong \mathbb{Z}/p_i^{m_{i1}}\mathbb{Z} \times \mathbb{Z}/p_i^{m_{i2}}\mathbb{Z} \times \cdots \mathbb{Z}/p_i^{m_{is_i}}\mathbb{Z}$$

for some $s_i \ge 1$ and $m_{i1} \ge m_{i2} \ge \cdots \ge m_{is_i} \ge 1$. Let $k = \max\{s_i \mid 1 \le i \le t\}$, set $m_{ij} = 0$ if $j > s_i$, and let

$$d_j = \prod_{i=1}^{t} p_i^{m_{ij}}$$

for each $1 \le j \le k$. Then $d_1, d_2, \dots, d_k \ge 2$, and d_{i+1} divides d_i for each $1 \le i \le k-1$, as desired. Moreover, Theorem 2.5.17 implies that

$$\mathbb{Z}/d_j\mathbb{Z} \cong (\mathbb{Z}/p_1^{m_{1j}}\mathbb{Z}) \times (\mathbb{Z}/p_2^{m_{2j}}\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p_t^{m_{tj}}\mathbb{Z}),$$

which yields the desired decomposition of G by gathering terms and applying these isomorphisms.

Finally, we address uniqueness of the latter two decompositions. First, we claim that in any decomposition of G (without restriction on the d_i), we must have the same r. For this, let

$$G_{\text{tor}} = \{a \in G \mid na = 0 \text{ for some } n \ge 1\}.$$

If we have written

$$G = \mathbb{Z}^r \times (\mathbb{Z}/d_1\mathbb{Z}) \times (\mathbb{Z}/d_2\mathbb{Z}) \times \cdots \times (\mathbb{Z}/d_k\mathbb{Z})$$

for some $r, k \ge 0$ and $d_1, d_2, \dots, d_k \ge 2$, then

$$G_{\text{tor}} = \{0\} \times (\mathbb{Z}/d_1\mathbb{Z}) \times (\mathbb{Z}/d_2\mathbb{Z}) \times \cdots \times (\mathbb{Z}/d_k\mathbb{Z}),$$

and $G/G_{\text{tor}} \cong \mathbb{Z}^r$. But $\mathbb{Z}^r \ncong \mathbb{Z}^s$ for $s \ne r$, so the r in the decomposition must be unique. Moreover, if $G_{\text{tor}} = \{0\}$, which is to say that $|G_{\text{tor}}| = 1$, then uniqueness of the decomposition is simply that r is unique such that $G \cong \mathbb{Z}^r$, which we have just proven.

Now, suppose $k \ge 1$ and we have chosen the d_i either to be prime powers, which we list in descending order, or such that d_{i+1} divides d_i for all $1 \le i \le k-1$ (so also in descending order). In the former case, d_1 is the largest order of any element of prime power order in G, and in the latter, d_1 is the exponent of G. Therefore, if we have a second decomposition,

$$G = \mathbb{Z}^r \times (\mathbb{Z}/d_1'\mathbb{Z}) \times (\mathbb{Z}/d_2'\mathbb{Z}) \times \cdots (\mathbb{Z}/d_{k'}'\mathbb{Z}),$$

written in the same form as the first, then we must have $d_1 = d_1'$. Take the quotient, therefore, by the subgroup $\mathbb{Z}^r \times \mathbb{Z}/d_1\mathbb{Z} \times \{0\} \times \{0\}$ in each decomposition. Then we have

$$\mathbb{Z}/d_2\mathbb{Z} \times \cdots \times \mathbb{Z}/d_k\mathbb{Z} \cong \mathbb{Z}/d_2'\mathbb{Z} \times \cdots \times \mathbb{Z}/d_{k'}'\mathbb{Z}.$$

By induction on the order of G_{tor}, assuming that we have proven the uniqueness for all G' with $|G'_{\text{tor}}| < |G_{\text{tor}}|$, we must have $k = k'$ and $d_i = d_i'$ for all $2 \le i \le k$, proving uniqueness. □

In the course of the proof of , we considered a subgroup of an abelian group G that we denote G_{tor}. We give it a formal definition.

DEFINITION 4.5.5. Let G be an abelian group.

a. The *torsion subgroup* of G is the subgroup

$$G_{\text{tor}} = \{a \in G \mid na = 0 \text{ for some } n \geq 1\}$$

of G.

b. An element of G_{tor} is called a *torsion element* of G.

REMARK 4.5.6. If G is a finite abelian group, then $G = G_{\text{tor}}$.

We leave the proof of the following direct corollary of Theorem 4.5.4 to the reader.

COROLLARY 4.5.7. *If G is a finitely generated abelian group, then $G \cong \mathbb{Z}^r \times G_{\text{tor}}$ for some $r \geq 0$. Moreover, G_{tor} is a finite abelian group, and it is isomorphic to a direct product of cyclic groups.*

4.6. Group actions on sets

DEFINITION 4.6.1. An *action* of a group G on a set X is an operation

$$\star\colon G \times X \to X$$

satisfying the following properties

i. $e \star x = x$ for all $x \in X$,

ii. $a \star (b \star x) = (ab) \star x$ for all $a, b \in G$ and $x \in X$.

We then say that G *acts* on X and that the operation $\star$ is a *G-action*, and we refer to X as a *G-set*.

REMARK 4.6.2. As with binary operations, we typically denote "$\star$" more simply by "$\cdot$".

EXAMPLES 4.6.3.

a. The symmetric group S_X acts on X. In particular, S_n acts on $X_n = \{1, 2, \dots, n\}$.

b. The group of isometries of $\mathbb{R}^n$ acts on $\mathbb{R}^n$.

c. The wallpaper group of a tiling of the plane acts on $\mathbb{R}^2$.

d. For $n \geq 3$, the group D_n acts on the set of vertices of the regular n-gon of which it is the symmetry group, as well as the set of its edges.

e. The group $\mathrm{GL}_n(\mathbb{R})$ acts on $\mathbb{R}^n$ by left multiplication of column vectors.

Here are a couple of more abstract examples.

EXAMPLES 4.6.4.

a. A group G acts on itself by left multiplication: $a \star x = ax$ for $a, x \in G$.

b. A group G acts on itself by conjugation: $a \star x = axa^{-1}$ for $a, x \in G$.

REMARK 4.6.5. A group G does not act on itself by right multiplication. If we defined $a \star x = xa$, then

$$a \star (b \star x) = a \star (xb) = (xb)a = x(ba),$$

while $(ab) \star x = x(ab)$. The action by right multiplication is an example of what is known as a right action (as opposed to a usual, or left, action).

DEFINITION 4.6.6. We say that an action of a group G on a set X is *transitive* if for every $x, y \in X$, there exists $a \in G$ with $ax = y$. We then say that G acts *transitively* on X.

EXAMPLES 4.6.7.

a. The group S_X acts transitively on X.

b. The group D_n acts transitively on the set of vertices of a regular n-gon, as well as the set of edges.

c. The group of isometries of $\mathbb{R}^n$ acts transitively on $\mathbb{R}^n$.

d. The group $\mathrm{GL}_n(\mathbb{R})$ does not act transitively on $\mathbb{R}^n$, as an invertible matrix times a nonzero vector is always nonzero.

e. Any group G acts transitively on itself by left multiplication. This is simply the cancellation theorem: if $x, y \in G$, then $a = yx^{-1}$ satisfies $ax = y$.

f. The action of G on itself by conjugation is not transitive if G is nontrivial. For example, the identity element is not a conjugate of any other element.

g. The group S_X acts on the power set of X,

$$\sigma \star Y = \{\sigma(y) \mid y \in Y\}$$

for $Y \subseteq X$, but this action is not transitive if X is nonempty. For instance, $\sigma \star Y$ always has the same cardinality as Y, so it cannot be the empty set if Y is nonempty.

h. The group G acts on the set G/H of left cosets a subgroup H of G by left multiplication:

$$a \cdot bH = abH,$$

and this is a transitive action.

DEFINITION 4.6.8. Let G be a group and X be a G-set. The *orbit* of $x \in X$ is the set

$$G \cdot x = \{gx \mid g \in G\}.$$

REMARK 4.6.9. Recall that the orbit of $x \in X_n$ under $\sigma \in S_n$ was defined as

$$O_\sigma(x) = \{\sigma^i x \mid i \in \mathbb{Z}\},$$

and we can reinterpret this orbit as the orbit $\langle \sigma \rangle \cdot x$.

The following lemma is nearly immediate.

LEMMA 4.6.10. *An action of a group G on a set X is transitive if and only if $G \cdot x = X$ for every (equivalently, some) $x \in X$.*

EXAMPLES 4.6.11.

a. Since S_X acts transitively on X, we have that $S_X \cdot x = X$ for every $x \in X$.

b. Consider the action of S_n on the power set of X_n. We have

$$S_n \cdot \{1, 2, \ldots, k\} = \{Y \subset X_n \mid |Y| = k\}.$$

c. The orbit of $v \in \mathbb{R}^n$ under $\mathrm{GL}_n(\mathbb{R})$ is $\mathbb{R}^n - \{0\}$ if $v \neq 0$ and $\{0\}$ if $v = 0$.

d. The orbit of $x \in G$ under the action of G on itself by conjugation is the conjugacy class C_x of x.

e. The orbit of $H \leqslant G$ under the action of G of its set of subgroups by conjugation is the set of all conjugate subgroups to G:

$$\{aHa^{-1} \mid a \in G\}.$$

We remark that the property of being in the same orbit is an equivalence relation on a G-set X, and therefore we obtain a partition of X as a disjoint union of its orbits.

PROPOSITION 4.6.12. *Let X be a G-set. The relation $x \sim_G y$ if and only if $G \cdot x = G \cdot y$ for $x, y \in X$ is an equivalence relation on G, and the equivalence class of $x \in X$ under $\sim_G$ is the orbit Gx of x. Therefore, G is the disjoint union of its distinct orbits.*

PROOF. That $\sim_G$ is an equivalence relation is checked immediately. We remark that for $x, y \in X$, we have $G \cdot x = G \cdot y$ if and only if $y \in G \cdot x$, since $ax = by$ for some $a, b \in G$ if and only if $cx = y$ for some $c \in G$ (that c being $b^{-1}a$). Therefore, the equivalence class of $x \in X$ is exactly the orbit $G \cdot x$, and the final statement is just Lemma 1.2.10. □

DEFINITION 4.6.13. Let X be a G-set for some group G. Let $a \in G$ and $x \in X$. We say that a *fixes* x if $ax = x$.

DEFINITION 4.6.14. We say that an action of a group G on a set X is *faithful* if the only element $a \in G$ that fixes all $x \in X$ is the identity element. We then say that G acts *faithfully* on X, and X is a *faithful* G-set.

In other words, G acts faithfully on X if $ax = x$ for all $x \in X$ implies $a = e$.

EXAMPLES 4.6.15.

a. The group S_X acts faithfully on X, since a nontrivial permutation of X does not fix every element of X.

b. For $n \geq 3$, the group D_n acts faithfully on the set of vertices of the regular polygon, as well as the set of edges.

c. The group G acts on itself faithfully by left multiplication, since if $ax = x$ for any $x \in G$, then $a = e$.

d. The action of a group G on itself by conjugation is faithful if and only if the group has trivial center. To see this, note that

$$Z(G) = \{a \in G \mid axa^{-1} = x \text{ for all } x \in G\}.$$

DEFINITION 4.6.16. Let G be a group and X be a G-set. The *stabilizer*, or *isotropy subgroup*, of G_x of an element $x \in X$ is the set of elements of G that fix x. That is, we have

$$G_x = \{a \in G \mid ax = x\}.$$

REMARK 4.6.17. The stabilizer G_x is indeed a subgroup of G, since $e \in G_x$, and for $a, b \in G_x$, we have $(ab)x = a(bx) = ax = x$, so $ab \in G_x$, while

$$a^{-1}x = a^{-1}(ax) = ex = x,$$

so $a^{-1} \in G_x$.

LEMMA 4.6.18. *A group G acts faithfully on a set X if and only if*

$$\bigcap_{x \in X} G_x = \{e\}.$$

PROOF. We have $a \in G_x$ if and only if $ax = x$. Thus $a \in G_x$ for all $x \in X$ if and only if $ax = x$ for all $x \in X$, and the action of G on X is not faithful if and only if the latter occurs for some $a \in G$ with $a \neq e$. So, $\bigcap_{x \in X} G_x$ contains a non-identity element if and only if G acts non-faithfully on X. □

EXAMPLES 4.6.19. We give some examples of stabilizers.

a. The stabilizer of n under the action of S_n is the image of S_{n-1} under the homomorphism $\iota\colon S_{n-1} \to S_n$ of Example 2.10.20.

b. The stabilizer of a vertex under the action of D_n on a regular n-gon consists exactly of the subgroup of order 2 generated by the unique reflection in D_n for which the line of reflection passes through the vertex.

c. The stabilizer of 4 under the action of $\langle \sigma \rangle \leqslant S_5$, where $\sigma = (1\ 2\ 3)(4\ 5)$, is $\langle \sigma^2 \rangle$.

d. The stabilizer of $x \in G$ under the action of G on itself by left multiplication is trivial:

$$G_x = \{a \in G \mid ax = x\} = \{e\}$$

The following definition gives an interesting class of examples of stabilizers.

DEFINITION 4.6.20. The stabilizer of $x \in G$ under the action of G on itself by conjugation is the subgroup of elements in G that commute with x and is known as the *centralizer* Z_x of x

$$Z_x = \{a \in G \mid ax = xa\}.$$

EXAMPLE 4.6.21. The centralizer $Z_{(1\ 2\ 3)}$ in S_5 is

$$Z_{(1\ 2\ 3)} = \langle (1\ 2\ 3), (4\ 5) \rangle.$$

We end by comparing orbits and stabilizers.

THEOREM 4.6.22. *Let X be a G-set, and let $x \in X$. Then there is a bijection*

$$\psi_x\colon G/G_x \to G \cdot x,$$

given by $\psi_x(aG_x) = ax$ for any $a \in G$.

PROOF. First, we note that ψ_x is well-defined, since if $b \in aG_x$, then $b = ag$ for some $g \in G_x$, and

$$\psi_x(bG_x) = bx = agx = a(gx) = ax = \psi_x(aG_x).$$

Moreover, it is one-to-one since, if $ax = bx$, then $x = a^{-1}bx$, so $a^{-1}b \in G_x$, and therefore $aG_x = bG_x$. Finally, it is onto by definition. □

COROLLARY 4.6.23. *If G is a finite group, then every element $x \in X$ has a finite orbit, and*

$$|G \cdot x| = [G : G_x].$$

In particular, the number of elements in the orbit of x divides $|G|$.

EXAMPLE 4.6.24. The centralizer of $(1\ 2\ 3)$ in S_5 has order 6, while the orbit of $(1\ 2\ 3)$ is the set of 3-cycles in S_5, of which there are 20, and we note that $|S_5| = 20 \cdot 6$.

We find an application in the class equation.

PROPOSITION 4.6.25 (The class equation). *Let G be a finite group. Then*

$$|G| = |Z(G)| + \sum_{x \in X} [G : Z_x],$$

where X is set of representatives of the conjugacy classes in G with more than one element.

PROOF. By Corollary 4.6.23, we have that $[G : Z_x] = |C_x|$ for $x \in X$. Moreover, $|C_x| = 1$ if $x \in Z(G)$. The equality we wish to prove is therefore reduced to the known fact that $|G|$ is the sum of the orders of its distinct conjugacy classes. □

Here is one application.

PROPOSITION 4.6.26. *The group A_5 is simple.*

PROOF. Any normal subgroup of a group is a disjoint union of conjugacy classes in that group including the conjugacy class $\{e\}$. Let us determine the conjugacy classes in A_5. The conjugacy classes in S_5 of nontrivial elements in A_5 are the products of 2 transpositions, the 3-cycles, and the 5-cycles. The centralizer of $(1\ 2\ 3)$ in A_5 is the group $\langle (1\ 2\ 3) \rangle$, so $|C_{(1\ 2\ 3)}| = \frac{60}{3} = 20$, and $C_{(1\ 2\ 3)}$ is thus the set of 3-cycles. The centralizer of $(1\ 2)(3\ 4)$ is $\langle (1\ 2)(3\ 4), (1\ 3)(2\ 4) \rangle$, so $|C_{(1\ 2)(3\ 4)}| = 15$, and $C_{(1\ 2)(3\ 4)}$ is thus the set of products of two transpositions. The centralizer of $(1\ 2\ 3\ 4\ 5)$ is $\langle (1\ 2\ 3\ 4\ 5) \rangle$, so $|C_{(1\ 2\ 3\ 4\ 5)}| = 12$ and so there are two conjugacy classes of 5 cycles, each with 12 elements. We then have that the distinct conjugacy classes of elements in A_5 have 1, 12, 12, 15, and 20 elements. Aside from 1 and 60, no sum of these numbers including 1 divides 60, so A_5 has no nontrivial, improper normal subgroups. □

4.7. Permutation representations

In this brief section, we give a characterization of group actions on sets as permutation representations.

THEOREM 4.7.1. *Let X be a G-set, and define $\sigma_a \colon X \to X$ by $\sigma_a(x) = ax$ for any $a \in G$. Then $\sigma_a \in S_X$, and the map*

$$\rho \colon G \to S_X$$

such that $\rho(a) = \sigma_a$ for all $a \in G$ is a homomorphism. Conversely, if $\rho \colon G \to S_X$ is a homomorphism, then the operation defined by

$$a \star x = \rho(a)(x)$$

on $a \in G$ and $x \in G$ is a G-action.

PROOF. We have

$$(\sigma_{a^{-1}} \circ \sigma_a)(x) = x = (\sigma_a \circ \sigma_{a^{-1}})(x),$$

so σ_a is a bijection, which is to say $\sigma_a \in S_X$. For $a, b \in G$ and $x \in X$, we have

$$\rho(ab)(x) = \sigma_{ab}(x) = (ab)x = a(bx) = \sigma_a(\sigma_b(x)) = (\sigma_a \circ \sigma_b)(x) = (\rho(a) \circ \rho(b))(x),$$

so ρ is a homomorphism.

Conversely, given ρ, we check that $e \star x = \rho(e)(x) = \mathrm{id}_X(x) = x$, while

$$a \star (b \star x) = \rho(a)(\rho(b)(x)) = \rho(ab)(x) = (ab) \star x,$$

so $\star$ is a G-action. □

DEFINITION 4.7.2. If X is a G-set, then the homomorphism ρ_X associated to X by Theorem 4.7.1 is called its *permutation representation.*

EXAMPLE 4.7.3. The action of S_X on X gives rise to a permutation representation $\rho\colon S_X \to S_X$ satisfying $\rho(\sigma)(x) = \sigma(x)$ for all $x \in X$. In other words, we have $\rho = \mathrm{id}_{S_X}$.

EXAMPLE 4.7.4. Consider the action of $\mathrm{GL}_n(\mathbb{R})$ on $\mathbb{R}^n$. The permutation representation

$$\rho\colon \mathrm{GL}_n(\mathbb{R}) \to S_{\mathbb{R}^n}$$

takes $A \in \mathrm{GL}_n(\mathbb{R})$ to a map $T_A\colon \mathbb{R}^n \to \mathbb{R}^n$ that satisfies $T_A(v) = Av$. In other words, the image of ρ is the subgroup of $S_{\mathbb{R}^n}$ consisting of invertible linear transformations.

The following lemma is almost immediate.

LEMMA 4.7.5. *A group G acts faithfully on a set X if and only if ρ_X is injective. In fact, the kernel of ρ_X is the intersection of the stabilizers G_x over all $x \in X$.*

PROOF. We have $\rho_X(a) = \mathrm{id}_X$ if and only if $ax = \rho_X(a)(x) = x$ for all $x \in X$. □

EXAMPLE 4.7.6. The permutation representation attached to G acting on itself by conjugation is a homomorphism $\gamma\colon G \to S_G$ given by $a \mapsto \gamma_a$, and its image is the inner automorphism group, a subgroup of $\mathrm{Aut}(G) \leqslant S_G$. The kernel of γ is $Z(G)$.

We now prove Cayley's theorem, which tells us that every group is a subgroup of a symmetric group.

THEOREM 4.7.7 (Cayley). *Every group G is isomorphic to a subgroup of S_G.*

PROOF. Consider the permutation representation $\rho_G\colon G \to S_G$ associated to the action of G on itself by left multiplication. By Example 4.6.4, ρ_G is injective, and therefore, G is isomorphic to $\mathrm{im}\,\rho_G \leqslant S_G$. □

REMARK 4.7.8. Note that if G has order n, Cayley's theorem tells us that G is isomorphic to a subgroup of S_n. However, this is not always the smallest permutation group in which it is contained. For example, the action of D_n on its vertices is faithful, and so a choice of numbering of these vertices identifies D_n with a subgroup of S_n, as opposed to S_{2n}. Even more simply, S_n is obviously a subgroup of itself, and not just isomorphic to a subgroup of $S_{n!}$.

4.8. Burnside's formula

NOTATION 4.8.1. Let X be a G-set, and let S be a subset of G. Then we set

$$X^S = \{x \in X \mid ax = x \text{ for all } a \in S\}.$$

If $S = \{a\}$ for some $a \in X$, we sometimes write X^a for $X^{\{a\}}$.

REMARK 4.8.2. For $x \in X$ and $a \in G$, where X is a G-set, the statement that $ax = x$ is equivalent both to $a \in G_x$ and to $x \in X^a$.

EXAMPLES 4.8.3.

a. Take $\sigma = (1\ 3)(2\ 5) \in S_6$, and let $X = X_6$. We have $X^\sigma = \{4,6\}$.

b. Let X be the D_n-set that is the set of vertices of the regular n-gon inscribed on the unit circle in $\mathbb{R}^2$ with a vertex at $(1,0)$. Then $X^s = \{(1,0)\}$ if n is odd and $X^s = \{(-1,0),(0,1)\}$ if n is even, while $X^r = \varnothing$.

c. Let $X = \mathbb{R}^n$, and let $A \in \mathrm{GL}_n(\mathbb{R})$. Then $X^A = \{v \in \mathbb{R}^n \mid Av = v\}$ is the eigenspace of A with eigenvalue 1 (so $\{0\}$ if 1 is not an eigenvalue). We have $X^{\mathrm{GL}_n(\mathbb{R})} = \{0\}$.

We now state Burnside's formula.

THEOREM 4.8.4 (Burnside). *Let G be a finite group, and let X be a finite G-set. Let r be the number of distinct orbits in X under G. Then*

$$r = \frac{1}{|G|}\sum_{a\in G}|X^a|.$$

PROOF. We will count the set of pairs

$$S = \{(a,x) \mid a \in G, x \in X, ax = x\} \subseteq G \times X$$

in two different ways. First, note that

$$S = \coprod_{a\in G}\{(a,x) \mid x \in X^a\},$$

so we have

$$|S| = \sum_{a\in G}|X^a|.$$

On the other hand, note that

$$S = \coprod_{x\in X}\{(a,x) \mid a \in G_x\},$$

so we have

$$|S| = \sum_{x\in X}|G_x| = \sum_{x\in X}\frac{|G|}{|G\cdot x|},$$

the latter statement being Corollary 4.6.23 and Lagrange's theorem. If $\mathscr{O}$ is an orbit in X, then it is the orbit of all $x \in \mathscr{O}$, so we have

$$\sum_{x\in\mathscr{O}}\frac{|G|}{|G\cdot x|} = |G|\sum_{x\in\mathscr{O}}\frac{1}{|\mathscr{O}|} = |G|.$$

Since X is the disjoint union of its orbits $\mathscr{O}$ and there are r orbits, we obtain $|S| = r|G|$, and therefore we have

$$r|G| = \sum_{a\in G}|X^a|,$$

as desired. □

Burnside's formula has an amusing use in certain problems involving counting.

EXAMPLE 4.8.5. Suppose we have a table with which is a regular octagon, with one chair placed at each side. Let us consider two seatings of eight people at the table to be the equivalent if and only if every person has the same two neighbors under both seatings (though possibly on different sides). We can ask: how many equivalence classes of seatings are there?

Let X denote the set of all seatings, so $|X| = 8!$. Two seatings are equivalent if and only if there is an element of D_8 that takes the positions of the people under one seating to the their positions under the other. In other words, the seatings are in one-to-one correspondence with the orbits under the action of D_8 on X. We note that $X^e = X$, and $X^a = \varnothing$ if $a \neq e$ since any nontrivial element of D_8 will change the position of at least one person (in fact, at least six people). Applying Burnside's formula, we have that the number of equivalence classes r of seatings is $8!/16 = 2520$.

EXAMPLE 4.8.6. How many different ways are there to color the faces of a cube either red or blue that actually look different? (Here: two colorings are the same if one is a rotation of another.)

Let X denote the set of all colorings, so $|X| = 2^6 = 64$. The group of rotations (orientation-preserving isometries) of a die has 5 types of elements: the identity, 6 rotations of order 4 through the centers of opposite faces, 3 rotations of order 2 of the same form, 6 rotations of order 2 through the centers of opposite edges, and 8 rotations of order 3 through the centers of opposite vertices. Respectively, these elements a have $|X^a| = 64, 8, 16, 8$, and 4. We then have

$$r = \frac{1}{24}(64 + 6\cdot 8 + 3\cdot 16 + 6\cdot 8 + 8\cdot 4) = \frac{240}{24} = 10.$$

This means there are exactly 10 different-looking colorings. Note that, as is often the case with these sorts of problems, it would have been easier to simply count them directly.

4.9. p-groups

DEFINITION 4.9.1. A group G is said to be a *p-group* if every element of G is finite of order a power of p.

Note that we have already classified the finite abelian p-groups up to isomorphism.

EXAMPLE 4.9.2. The group D_4 is a nonabelian 2-group of order 8.

All finite groups of p-power order are clearly p-groups. We shall see that the converse is true as well. For this, we require the following useful lemma.

LEMMA 4.9.3. *Let G be a finite group of p-power order, and let X be a G-set. Then*

$$|X| \equiv |X^G| \bmod p.$$

PROOF. By Corollary 4.6.23, every orbit in X has order dividing $|G|$, hence a power of p. Note that the orbits of order 1 are exactly the $\{x\}$ with $x \in X^G$. On the other hand, the other orbits all have order divisible by p, so if Y is a set of representatives of the orbits of G, then it contains X^G, and we have

$$|X| = \sum_{y \in Y} |G \cdot y| \equiv \sum_{x \in X^G} |\{x\}| \bmod p,$$

□

In general, if the order of a finite group G is n, then while we know that every element of G has order dividing n, we do not have the converse (unless G is cyclic). On the other hand, Cauchy's theorem, which we now prove, tells us that G contains elements of every prime order dividing n.

THEOREM 4.9.4 (Cauchy). *Let p be a prime number, and let G be a finite group of order divisible by p. Then G contains an element of order p.*

PROOF. We consider the set

$$X = \{(a_1, a_2, \dots, a_p) \in G^p \mid a_1 a_2 \cdots a_p = e\}.$$

Note that if $(a_1, a_2, \dots, a_p) \in X$, then $a_1, a_2, \dots, a_{p-1} \in G$ can be chosen arbitrarily, and then $a_p = (a_1 a_2 \dots a_{p-1})^{-1}$ is determined by those a_i. It follows that $|X| = |G|^{p-1}$. Let $\tau = (1\ 2\ \dots\ p) \in S_p$. We let τ, and hence $\langle \tau \rangle$, act on X by

$$\tau \cdot (a_1, a_2, \dots, a_p) = (a_2, \dots, a_p, a_1).$$

Note that this is an action, as

$$(a_2 \cdots a_p) a_1 = e,$$

since a_1 being left inverse to $a_2 \cdots a_p$ implies that it is also right inverse to $a_2 \cdots a_p$. Then

$$X^{\langle \tau \rangle} = X^{\tau} = \{(a, a, \dots, a) \in G^p \mid a^p = e\}.$$

By Lemma 4.9.3, we have

$$|X| \equiv |X^{\langle \tau \rangle}| \bmod p.$$

Since p divides $|G|$, it divides $|X|$, and hence it divides the order of $X^{\langle \tau \rangle}$. But $X^{\langle \tau \rangle}$ is in bijection with the set of elements of G of order dividing p, and e is such an element. So, we must have at least p distinct elements in G of order dividing p, hence at least $p-1$ of order p. □

Cauchy's theorem has the above-mentioned corollary.

COROLLARY 4.9.5. *Every finite p-group has p-power order.*

PROOF. If G is a finite group and ℓ is a prime dividing $|G|$, then G has an element of order ℓ by Cauchy's theorem. So if G is a p-group, then by definition the only prime that can divide $|G|$ is p. □

The following result is very useful in the study of p-groups.

PROPOSITION 4.9.6. *The center $Z(G)$ of a nontrivial finite p-group is nontrivial.*

PROOF. Consider the action of G on itself by conjugation. The set of elements of G fixed by every element of G under conjugation is exactly the center of G. By Lemma 4.9.3, we therefore have that $|G| \equiv |Z(G)| \bmod p$. Since $|G|$ is a nontrivial power of p, this means that $Z(G)$ is not the trivial subgroup. □

We give an application of Proposition 4.9.6 to the study of the structure of p-groups of order p^2.

THEOREM 4.9.7. *Every group of order p^2, where p is a prime, is abelian.*

PROOF. Let G be a group of order p^2. By Proposition 4.9.6, we have that $Z(G)$ is nontrivial, so has either order p or p^2. We must show that it is the latter, since $Z(G) = G$ if and only if G is abelian. So, suppose by way of contradiction that $|Z(G)| = p$, and let $b \in G$ be an element that is not in the center of G. Then $H = Z(G)\langle b\rangle$ has order greater than p, hence is all of G. But H is abelian, since b commutes with every element of $Z(G)$ and certainly every element of $Z(G)$ commutes with itself. So, $G = H$ is abelian as well, contradicting $|Z(G)| = p$. □

In particular, this tells us that there are only two isomorphism classes of groups of order p^2, those of $\mathbb{Z}/p^2\mathbb{Z}$ and $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$.

4.10. The Sylow theorems

DEFINITION 4.10.1. Let G be a group, and let p be a prime number.

a. A subgroup of G is called a *p-subgroup* if it is a p-group.

b. A p-subgroup of G is called a *Sylow p-subgroup* of G if it is not properly contained in any p-subgroup of G.

REMARK 4.10.2. If G is a finite group and p^n is the largest power of p dividing $|G|$, then every p-subgroup of G has order dividing p^n.

EXAMPLES 4.10.3.

a. In S_5, the Sylow 5-subgroups are the subgroups generated by the 5-cycles, the Sylow 3-subgroups are the subgroups generated by the 3-cycles, and the Sylow 2-subgroups are the subgroups of order 8 the form

$$\langle (a\ b\ c\ d), (a\ c)\rangle$$

with a, b, c, d distinct elements of X_5.

b. If $n \geq 3$ and $n = 2^k m$ with m odd, then every Sylow 2-subgroup of D_n has the form $D_n = \langle r^i s, r^m\rangle$ for some $0 \leq i < m$.

The Sylow theorems, which we now state in one compact result that we refer to as Sylow's theorem, constitute an extremely useful tool for the study of finite groups.

THEOREM 4.10.4 (Sylow). *Let G be a finite group, let p be a prime number, and let n be exponent of the highest power of p dividing $|G|$. Then the following hold.*

a. Every Sylow p-subgroup of G has order p^n.

b. Every two Sylow p-subgroups of G are conjugate.

c. *The number of Sylow p-subgroups divides $|G|$ and is congruent to* 1 *modulo p.*

We defer the proof of the Sylow's theorem to below.

NOTATION 4.10.5. Let G be a finite group and p a prime number. We let $\mathrm{Syl}_p(G)$ denote the set of Sylow p-subgroups of G, and we let $n_p(G) = |\mathrm{Syl}_p(G)|$.

The fact that $n_p(G)$ both divides $|G|$ and is congruent to 1 modulo P can be very useful in determining the possible isomorphism classes of groups of a given order.

EXAMPLE 4.10.6. It is easy to see from our description of the Sylow p-subgroups of S_5 that every two Sylow p-sugroups of S_5 are conjugate, as the elements generating such groups are conjugate. We have $n_5(S_5) = 6 \equiv 1 \bmod 5$. We also have $n_3(S_5) = 10 \equiv 1 \bmod 3$, and $n_2(S_5) = 15$, which is odd.

To understand the second part of Sylow's theorem, we introduce the concept of a normalizer.

DEFINITION 4.10.7. Let G be a group and H be a subgroup. The *normalizer* $N_G(H)$ of H in G is the subgroup

$$N_G(H) = \{a \in G \mid aHa^{-1} = H\}$$

of G.

REMARKS 4.10.8.

a. By definition, $N_G(H)$ is the stabilizer of H under the action of G on its set of subgroups by conjugation, so in particular is a subgroup.

b. We have $H \trianglelefteq N_G(H)$ and $N_G(H)$ is the largest subgroup of G in which H is normal.

c. We have $N_G(H) = G$ if and only if $H \trianglelefteq G$.

EXAMPLES 4.10.9.

a. We have $N_{D_n}(\langle r \rangle) = D_n$ and $N_{D_n}(\langle s \rangle) = \langle s \rangle$ if n is odd and $\langle s, r^{n/2} \rangle$ if n is even.

b. The normalizer of $\langle (1\ 2\ 3\ 4) \rangle$ in S_5 is $\langle (1\ 2\ 3\ 4), (1\ 3) \rangle$.

The following lemma is crucial for proving the first part of Sylow's theorem.

LEMMA 4.10.10. *Suppose that G is a finite group, and let H be a subgroup of G of order a power of a prime p. Then*

$$[G : H] \equiv [N_G(H) : H] \bmod p.$$

PROOF. Let $\mathscr{L} = G/H$. Then H acts on $\mathscr{L}$ by left multiplication: $h \cdot (aH) = haH$ for $h \in H$, $a \in G$. We have that $aH \in \mathscr{L}^H$ if and only if $h(aH) = aH$ for every $h \in H$, which is to say that $aha^{-1} \in H$ for every $h \in H$, which means exactly that $aHa^{-1} = H$. In other words,

$$\mathscr{L}^H = \{aH \mid a \in N_G(H)\} = N_G(H)/H.$$

By Lemma 4.9.3, we have that

$$|\mathscr{L}| \equiv |\mathscr{L}^H| \bmod p,$$

which is exactly the statement of the lemma. □

Let us fix a prime p throughout the rest of this section. We prove a strengthening of the first part of Sylow's theorem.

THEOREM 4.10.11 (First Sylow theorem). *Let G be a group, and let n be the exponent of the highest power of p dividing G. Every subgroup of G of order p^k with $k < n$ is a normal subgroup of a subgroup of G of order p^{k+1}.*

PROOF. Suppose that $H < G$ has order p^k. By Lemma 4.10.10, its index in its normalizer is congruent to $[G : H]$ modulo p, so is divisible by p. But then $N_G(H)/H$ has order divisible by p, and so by Cayley's theorem there exists a subgroup of it of order p. By Proposition 2.13.10, there then exists a subgroup K of $N_G(H)$ in which H is normal and such that $|K/H| = p$. Lagrange's theorem then implies that $|K| = p^{k+1}$, as desired. □

Theorem 4.10.11 tells us, in particular, that every p-subgroup of G of order less than p^n is not maximal, so part a of Sylow's theorem holds. In fact, recursion tells us that:

COROLLARY 4.10.12. *Every p-subgroup of a finite group G is contained in a Sylow p-subgroup of order p^n, where n is the exponent of the highest power of p dividing G. In particular, every Sylow p-subgroup has order p^n.*

The first Sylow theorem also has the following simple corollary.

COROLLARY 4.10.13. *Let G be a group, and let n be the exponent of the highest power of p dividing G. The G has subgroups of order p^k for every $1 \le k \le n$.*

PROOF. Suppose without loss of generality that p divides $|G|$. By Cauchy's theorem, G has an element of order p, so it has a subgroup of order p. By recursion, the first Sylow theorem then tells us that G has subgroups of every p-power order dividing G. □

We next prove the second part of Sylow's theorem, which we state as a separate result.

THEOREM 4.10.14 (Second Sylow theorem). *If P and Q are Sylow p-subgroups of a finite group G for some prime p, then P and Q are conjugate subgroups of G.*

PROOF. We consider the action of Q on the set of left cosets G/P via $h \cdot aP = (ha)P$ for $h \in Q$ and $a \in G$. By Lemma 4.9.3, we have that $|(G/P)^Q| \equiv |G/P| \bmod p$. Since p does not divide $|G/P|$, we therefore have that p does not divide $|(G/P)^Q|$. In particular, there exists an element $bP \in (G/P)^Q$. Since $hbP = bP$ for all $h \in Q$, we have $b^{-1}hb \in P$ for all $h \in Q$, so $b^{-1}Qb \leqslant P$. Since P and Q have the same order, we therefore have that $Q = bPb^{-1}$ is a conjugate of P in G. □

COROLLARY 4.10.15. *Suppose that G is a finite group, and let P be a Sylow p-subgroup of G. Then P is normal in G if and only if $n_p(G) = 1$.*

PROOF. We know that every conjugate of a Sylow p-subgroup is a Sylow p-subgroup, as it has the same order, so it is an immediate corollary of Theorem 4.10.14 that P is normal in G if and only if $n_p(G) = 1$. □

The second Sylow theorem also has, after a short argument, the following consequence.

PROPOSITION 4.10.16. *Let G be a finite group, let p be a prime number, and let P be a Sylow p-subgroup. Then $n_p(G) = [G : N_G(P)]$. In particular, the number of Sylow p-subgroups of G divides $|G|$.*

PROOF. Consider the action of G on the set $\mathrm{Syl}_p(G)$ of Sylow p-subgroups of G by conjugation By Theorem 4.10.4b, we have that $\mathrm{Syl}_p(G)$ has just one orbit under this action, which is all of $\mathrm{Syl}_p(G)$. Since the stabilizer of P is $N_G(P)$, the result follows from Corollary 4.6.23. □

We will require a special case of the following lemma.

LEMMA 4.10.17. *Let Q be a Sylow p-subgroup and P be a p-subgroup of a finite group G. Then $P \cap N_G(Q) = P \cap Q$.*

PROOF. Let $H = P \cap N_G(Q)$. We need only show that $H \leqslant Q$. Since $H \leqslant N_G(Q)$, we have that $hQh^{-1} = Q$ for all $h \in H$, so $HQ = QH$, and therefore HQ is a subgroup of G with Q as a normal subgroup. The second isomorphism theorem implies that

$$|HQ| = \frac{|H||Q|}{|H \cap Q|},$$

and yields in particular that HQ is a p-subgroup of G. On the other hand, Q is a Sylow p-subgroup of G, so HQ cannot be larger, and therefore must equal Q. Thus, we have the required containment $H \leqslant Q$. □

Finally, we prove the third part of Sylow's theorem.

THEOREM 4.10.18 (Third Sylow theorem). *The number $n_p(G)$ of Sylow p-subgroups of a finite group G divides $|G|$ and is congruent to 1 modulo p.*

PROOF. The first part is just Proposition 4.10.16 and Lagrange's theorem. For the second part, we assume that p divides $|G|$, as the result is otherwise trivial. Let P be a Sylow p-subgroup of G, and let P act on $\mathrm{Syl}_p(G)$ by conjugation: if $Q \in \mathrm{Syl}_p(G)$ and $a \in P$, then a takes Q to aQa^{-1}. By Lemma 4.9.3, we then have

$$n_p(G) \equiv |\,\mathrm{Syl}_p(G)^P| \bmod p.$$

Let $Q \in \mathrm{Syl}_p(G)^P$, which tells us that $P \leqslant N_G(Q)$. By Lemma 4.10.17, we then have that

$$P = P \cap N_G(Q) \leqslant Q,$$

which forces $P = Q$ as P and Q have the same order by Corollary 4.10.12. Thus, we have that $|\,\mathrm{Syl}_p(G)^P| = 1$, and so $n_p(G) \equiv 1 \bmod p$. □

4.11. Applications of Sylow theory

We can use Sylow's theorem to classify, or simply to give information on, the structure of groups of a given order. For instance, Sylow's theorem can be used to show that there are no simple groups of certain small orders or of orders with certain sorts of prime factorizations, as we see in the following examples.

EXAMPLE 4.11.1. There are no simple groups of order 42. If G is a group of order 42 , then $n_7(G)$ divides 42 and is 1 modulo 7, which forces $n_7(G) = 1$. By Corollary 4.10.15, we have that the unique subgroup of G of order 7 is normal, so G is not simple.

EXAMPLE 4.11.2. Let G be a group of order 30. Suppose that $n_3(G) > 1$ and $n_5(G) > 1$. Then the third part of Sylow's theorem tells us that $n_5(G) = 6$ and $n_3(G) = 10$. Now, any Sylow 5-subgroup has order 5, hence is cyclic with 4 elements of order 5, and any two distinct Sylow 5-subgroups have trivial intersection. Therefore, G contains 24 elements of order 5. On the other hand, the same argument with 3 replacing 5 tells us that G contains 20 elements of order 3. This is clearly impossible. In particular, there are no simple groups of order 30.

EXAMPLE 4.11.3. There are no simple groups of order p^n, where p is a prime and $n \geq 2$. This follows from Corollary 4.10.13, which tells us that such a group has a subgroup of order p^{n-1} and the first Sylow theorem, which tells us that the subgroup is normal in a subgroup of order p^n, which is necessarily the whole group.

We can also study groups with orders having a particularly nice form. The following result is useful for that.

PROPOSITION 4.11.4. *Let G be a group. Suppose that H and K are normal subgroups of G with $HK = G$ and $H \cap K = \{e\}$. Then the function $\psi \colon H \times K \to G$ given by $\psi(h,k) = hk$ for $h \in K$ and $k \in K$ is an isomorphism.*

PROOF. Let $h \in H$ and $k \in K$. Then $[h,k]$ equals both $(hk^{-1}h^{-1})k^{-1}$, from which it is seen to be an element of K, as $K \lhd G$, and $h(kh^{-1}k^{-1})$, which is similarly seen to be an element of H. As $H \cap K = \{e\}$, we therefore have $[h,k] = e$, and therefore elements of H commute with elements of K. It follows that ψ as defined is a homomorphism. It is onto as $G = HK$ and one-to-one as $hk = e$ implies $h,k \in H \cap K$, so $(h,k) = (e,e)$. □

Proposition 4.11.4 has the following application in conjunction with Sylow's theorems.

THEOREM 4.11.5. *Suppose that p and q are prime numbers with $p < q$. Then every group of order pq has a normal subgroup of order q and is in fact cyclic if $q \not\equiv 1 \bmod p$.*

PROOF. Note that $n_q(G)$ divides p and is 1 modulo q, which forces $n_q(G) = 1$ since $p < q$. By Corollary 4.10.15, G has a unique, normal Sylow q-subgroup K of order q. On the other hand, $n_p(G)$ divides q and is 1 modulo p. Supposing that $q \not\equiv 1 \bmod p$. then we must have $n_p(G) = 1$ as q is a prime. Let H be the unique, normal subgroup of order p. Now, both H and K are cyclic, so let $h,k \in G$ with $H = \langle h \rangle$ and $K = \langle k \rangle$. By Proposition 4.11.4, we have that

$$G \cong H \times K \cong \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z} \cong \mathbb{Z}/pq\mathbb{Z},$$

the last step being the Chinese remainder theorem. □

We can use this to give a more complicated example of a proof that all groups of a given order are cyclic.

EXAMPLE 4.11.6. All groups of order 255 are cyclic. Since $255 = 3 \cdot 5 \cdot 17$, the structure theorem for finitely generated abelian groups tells us that every abelian group of order 255 is cyclic. So, we must show that every group G of order 255 is abelian. By the third Sylow theorem, $n_{17}(G) = 1$, so G has a unique subgroup N of order 17, which is normal. Then G/N has order 15 and so is cyclic by Theorem 4.11.5. By Theorem 4.2.8, the subgroup N must contain the commutator subgroup $[G,G]$ of N.

Again by the third Sylow theorem, we have either $n_3(G) = 1$ or $n_3(G) = 85$ and $n_5(G) = 1$ or $n_5(G) = 51$. If $n_3(G) = 85$, then G has at least 170 elements of order 3, and if $n_5(G) = 51$, then G has at least 204 elements of order 5. Clearly, both of these cannot hold at the same time, so either $n_3(G) = 1$ or $n_5(G) = 1$. But then G has either a normal subgroup of order 3 or a normal subgroup of order 5. Call this subgroup Q. Then G/Q has order $3 \cdot 17$ or $5 \cdot 17$, and in either case, Theorem 4.11.5 tells us that it is cyclic. As before, we then have that Q contains $[G,G]$, but $Q \cap N$ is trivial since Q and N have relatively prime order, so $[G,G] = \{e\}$, which is to say that G is abelian.

Let us also expand our study of groups of order 30.

EXAMPLE 4.11.7. Every group G of order 30 has a normal subgroup of order 5. To see this, let P be a subgroup of order 3 and Q a subgroup of order 5. By Example 4.11.2, either P or Q is normal, and therefore PQ is a subgroup of G of order 15. By Theorem 4.11.5, it is cyclic. We thus have that $|N_G(Q)|$ is either 15 or 30, so $n_G(Q) = 1$ or 2, and 2 is impossible by the third Sylow theorem. Thus, Q is normal.

We can also rule out a whole class of possible orders of simple groups with the following result.

PROPOSITION 4.11.8. *There are no simple groups of order p^2q, where p and q are distinct prime numbers.*

PROOF. Let G be a group of order p^2q. If $p > q$, then $n_p(G) = 1$ by the third Sylow theorem, so G has a normal Sylow p-subgroup (which is abelian of order p^2). If $q > p$, then $n_q(G) = 1$ or p^2. We need only check the latter case. In this case, $p^2 \equiv 1 \bmod q$, so q divides $p^2 - 1$, but it does not divide $p - 1$ as $q > p$, so q divides $p + 1$, which forces $p = 2$ and $q = 3$. Then G has order 12. Now, if $n_3(G) = 4$, a simple element count shows that one cannot have $n_2(G) = 3$, so G has a normal Sylow 2-subgroup. □

Another method for exhibiting the non-simplicity of groups of a given order comes from the use of permutation representations.

PROPOSITION 4.11.9. *Let G be a finite simple group of order properly divisible by p. Then G is isomorphic to a subgroup of S_d where $d = n_p(G)$. In particular, $|G|$ divides $d!$.*

PROOF. Let G act on the set $\mathscr{L}$ of left cosets of the normalizer $N_G(P)$ of some Sylow p-subgroup P of G by left multiplication. This is a transitive action, so the permutation representation $G \to S_{\mathscr{L}}$ is nontrivial, hence injective as G is simple. As $|\mathscr{L}| = d$ by the third Sylow theorem, we have $S_{\mathscr{L}} \cong S_d$, so G is isomorphic to a subgroup of S_d. □

We provide a couple of examples.

EXAMPLE 4.11.10. There are no simple groups G of order 160. That is, if G were such a group, then $n_2(G) = 5$ by the third Sylow theorem, and therefore G is isomorphic to a subgroup of S_5. But 120 does not divide 5!.

EXAMPLE 4.11.11. There are no simple groups of order $396 = 2^2 3^2 11$. If G were such a group, then $n_{11}(G) = 12$, and the normalizer of a Sylow 11-subgroup P in G has order $33 = \frac{396}{12}$ by the third Sylow theorem. It follows by Proposition 4.11.9 that G is isomorphic to and thus may be identified with a subgroup of S_{12}. By definition $N_G(P)$ is contained in $N_{S_{12}}(P)$. But P is generated by an 11-cycle, and the number of such Sylow 11-subgroups of S_{12} is easily counted to be $12 \cdot 9!$ (as there are $\frac{12!}{11}$ such cycles and 10 per subgroup), which again by the third Sylow theorem implies that the order $N_P(G)$ is 110, which is not a multiple of 33.

The following weakening of the second isomorphism theorem to allow arbitrary finite subgroups is a useful tool.

LEMMA 4.11.12. *Let H and K be finite subgroups of a group G. Then we have*

$$|HK| = \frac{|H||K|}{|H \cap K|}.$$

PROOF. By definition, we have $|HK| \le |H||K|$. If $h, h' \in H$ and $k, k' \in K$ are such that $hk = h'k'$, then setting $a = (h')^{-1}h = k'k^{-1}$, we have that $a \in H \cap K$. Moreover, note that $h' = ha^{-1}$ and $k' = ak$. Conversely, given $h \in H$, $k \in K$, and $a \in H \cap K$, then defining $h' = ha^{-1}$ and $k' = ak$, we see that $hk = h'k'$. Therefore, if we define an equivalence relation on the set $H \times K$ by $(h,k) \sim (h',k')$ if and only if $hk = h'k'$, the number of pairs in each equivalence class is $|H \cap K|$, and as a result there are $|H||K|/|H \cap K|$ equivalence classes. On the other hand, the number of equivalence classes is by definition $|HK|$, proving the desired equality. □

Let us apply Lemma 4.11.12 to an example.

EXAMPLE 4.11.13. There are no simple groups of order 48. Suppose G is a group of order 48. By the third Sylow theorem, we have $n_2(G) = 1$ or 3. If $n_2(G) = 1$, then the unique Sylow 2-subgroup is normal. If $n_2(G) = 3$, let H and K be distinct subgroups of G of order 16. Then the fact that $|HK| \le 48$ and Lemma 4.11.12 force $|H \cap K| = 8$. Then $H \cap K$ has index 2 in H and K, hence is normal in both, so its normalizer $N_G(H \cap K)$ contains HK, which has order 32, so equals G. But then $H \cap K$ is normal in G, so again G is not simple.

The latter example used a special case of the following, which tells us that a group G of order 48 with $n_2(G) = 3 \not\equiv 1 \bmod 4$ has two Sylow 2-subgroups with intersection of order 8.

PROPOSITION 4.11.14. *Let G be a finite group, and let n be the exponent of the highest power of p dividing $|G|$. Let P be a Sylow p-subgroup of G. Let $r \le n$ be a positive integer such that $|P \cap Q| \le p^{n-r}$ for every Sylow p-subgroup Q of G with $Q \ne P$. Then we have $n_p(G) \equiv 1 \bmod p^r$.*

PROOF. Let P be a Sylow p-subgroup, and consider the action of P on $\mathrm{Syl}_p(G)$ by conjugation. Let $Q \in \mathrm{Syl}_p(G)$ with $Q \ne P$. Lemma 4.10.17 tells us that the elements in P that fix Q under conjugation (i.e., the $a \in P$ such that $aQa^{-1} = Q$) are exactly those in $P \cap Q$. For i such that $p^i = [P : P \cap Q]$, this implies that there are exactly p^i conjugates of Q by elements of P, so the order of the P-orbit of Q is p^i. Under the assumption of the proposition, we have that $|P \cap Q|$ divides p^{n-r}, so $i \geq r$. Therefore, P-orbit of $\operatorname{Syl}_p(G)$ other than the singleton orbit $\{P\}$ has order divisible by p^r, which implies that $n_p(G) = |\operatorname{Syl}_p(G)| \equiv 1 \bmod p^r$. □

4.12. Simplicity of alternating groups

Before we proceed to simplicity, we first show that we can use group actions to give an alternate definition of the sign of a permutation (and therefore of alternating groups) that does not use the determinant map, which we did not define above.

PROPOSITION 4.12.1. *Let S_n act on the set X of polynomials $p = p(x_1, x_2, \dots, x_n)$ in n variables $x_1, x_2, \dots, x_n$ by*

$$\sigma \cdot p = p(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)})$$

for $\sigma \in S_n$ and $p \in X$. Take

$$\Delta = \prod_{1 \leq i < j \leq n} (x_j - x_i) \in X$$

The function $\varepsilon \colon S_n \to \{\pm 1\}$ given by

$$\sigma \cdot \Delta = \varepsilon(\sigma)\Delta$$

for all $\sigma \in S_n$ is equal to the homomorphism sign.

PROOF. For $\sigma, \tau \in S_n$ and $p \in X$, we have

$$\sigma\tau \cdot p = p(x_{\sigma(\tau(1))}, x_{\sigma(\tau(2))}, \dots, x_{\sigma(\tau(n))}) = \sigma \cdot p(x_{\tau(1)}, x_{\tau(2)}, \dots, x_{\tau(n)}) = \sigma \cdot (\tau \cdot p),$$

and clearly $e \cdot p = p$ for the identity e, so S_n acts on X. Note that since $\sigma \in S_n$ takes each term $x_j - x_i$ with $i < j$ to plus or minus another term of the same form in a one-to-one fashion (in that σ is one-to-one), we do indeed have $\sigma \cdot \Delta = \pm\Delta$. That ε is a homomorphism follows from the fact that S_n acts on X, since

$$\varepsilon(\sigma\tau)\Delta = \sigma\tau(\Delta) = \sigma(\varepsilon(\tau)\Delta) = \varepsilon(\tau)\sigma(\Delta) = \varepsilon(\sigma)\varepsilon(\tau)\Delta.$$

It remains only to check that $\varepsilon(\tau) = -1$ for any transposition $\tau = (k\,\ell)$ with $1 \leq k < \ell \leq n$. For this, note that $\tau(x_j - x_i) = x_j - x_i$ unless $\{i, j\} \cap \{k, l\} = \varnothing$. If $\{i, j\} = \{k, \ell\}$, then $\tau(x_\ell - x_k) = -(x_\ell - x_k)$. The remaining terms have the form $\pm(x_\ell - x_m)$ or $\pm(x_k - x_m)$ for some $m \neq k, \ell$. We consider these in pairs. If $m < k$, then we have

$$\tau((x_\ell - x_m)(x_k - x_m)) = (x_k - x_m)(x_\ell - x_m) = (x_\ell - x_m)(x_k - x_m).$$

If $m > \ell$, we have

$$\tau((x_m - x_k)(x_m - x_\ell)) = (x_m - x_\ell)(x_m - x_k) = (x_m - x_k)(x_m - x_\ell),$$

and if $k < m < \ell$, we have

$$\tau((x_m - x_k)(x_\ell - x_m)) = (x_m - x_\ell)(x_k - x_m) = (x_m - x_k)(x_\ell - x_m).$$

Therefore, the product of the contributions to $\varepsilon(\tau)$ from the various terms is -1, as required. □

We once again exhibit that A_5 is simple, and moreover, that it is the only simple subgroup of order 60, up to isomorphism.

LEMMA 4.12.2. *If G is a group of order* 60 *with $n_5(G) > 1$, then G is simple.*

PROOF. The assumption forces $n_5(G) = 6$ by the third Sylow theorem. So, the normalizer of any Sylow p-subgroup has order 10. Let N be a proper normal subgroup of G. If $5 \mid |N|$, then N contains a Sylow 5-subgroup of G and hence all Sylow 5-subgroups of G by the second Sylow theorem. But then N has at least $1 + 6 \cdot 4 = 25$ elements, so is of order 30. But N has a unique subgroup of order 5 by Example 4.11.7, which is normal in G by Lemma 4.3.15a and Lemma 4.3.16. This contradicts $n_5(G) = 6$. It follows that $5 \nmid |N|$. Now, if N has order 6 or 12, then again it has a normal Sylow subgroup which is then by the same reasoning itself normal in G. So, we may assume that $|N| \in \{2,3,4\}$. Then $|G/N| \in \{15,20,30\}$, and in all of these cases, G/N has a normal subgroup of order 5 by the third Sylow theorem and Example 4.11.7. But then G itself has a normal subgroup with order divisible by 5, which we have already shown is not the case. Thus N must be the trivial subgroup. □

As a corollary, we recover Proposition 4.6.26 that A_5 is simple. Let us prove that this is the only subgroup of order 60.

PROPOSITION 4.12.3. *The group A_5 is isomorphic to every simple group of order* 60.

PROOF. Let G be a simple group of order 60, which we know exists by Proposition 4.6.26. We show that G is isomorphic to A_5. From the third Sylow theorem, the possibilities for $n_2(G)$ are 3, 5, and 15, which is also the index of the normalizer N of a Sylow 2-subgroup P. Since G is not isomorphic to a subgroup of S_3, we can eliminate $n_2(G) = 3$.

If $n_2(G) = 5$, then G is isomorphic to a subgroup of S_5, so G may be identified with a normal subgroup of S_5 of index 2. It follows that $G \cap A_5$ is a normal subgroup of A_5 which is either A_5 or of index 2 in A_5. The latter being impossible by the simplicity of A_5, we must have that $G = A_5$.

Suppose now that $n_2(G) = 15$. Since $15 \not\equiv 1 \bmod 4$, Proposition 4.11.14 tells us that $|P \cap Q| = 2$ for some $Q \in \mathrm{Syl}_2(G)$ with $Q \neq P$. Set $M = N_G(P \cap Q)$, which is not G since G is simple. Since $|M|$ is a multiple of 4 that is greater than 8 by Lemma 4.11.12, we must have $|M| = 12$ or 20, from which it follows that M has index at most 5, and therefore G is isomorphic to a subgroup of S_5. The same argument as before would tell us that $G \cong A_5$, but note that we assumed $n_2(G) = 15$, so we reach a contradiction. □

We now prove that the alternating groups on at least 5 elements are simple.

THEOREM 4.12.4. *The groups A_n for $n \geq 5$ are simple.*

PROOF. We prove this by induction on $n \geq 5$, the case $n = 5$ having been proven in Proposition 4.6.26. Let $G = A_n$ for some $n \geq 6$. For any $i \in X_n$, the stabilizer G_i is isomorphic to A_{n-1}, which is simple by induction. Suppose that N is a nontrivial normal subgroup of G. If there exists $i \in X_n$ and $\tau \in N - \{e\}$ with $\tau(i) = i$, then $N \cap G_i$ is a nontrivial normal subgroup of G_i, and it follows that $G_i \leqslant N$ by the simplicity of G_i. For any $j \in X_n$, we can find $\sigma \in A_n$ with $\sigma(i) = j$, and then $G_j = \sigma G_i \sigma^{-1} \leqslant N$ by normality of N. As every element of A_n can be written as a product of an even number of transpositions, every element of A_n may be written as a product of products of two transpositions, and any product of two transpositions lies in G_j for some $j \in X_n$ since $n > 4$. Thus, we must have that $N = G = A_n$.

Now, we show that N must contain a permutation that fixes some element of X_n. Let $\tau \in N$. If $\tau(i) = \tau'(i)$ for any $\tau' \in N$ and $i \in X_n$, then $\tau'\tau^{-1}$ fixes i. If the cycle decomposition of τ contains a k-cycle with $k \geq 3$, say $(a_1\ a_2\ \dots\ a_k)$, then we may choose $\sigma \in A_n$ that fixes a_1 and a_2 but not a_3. If τ is a product of disjoint transpositions that does not fix any element, then write $\tau = (a_1a_2)(a_3a_4)\dots(a_{m-1}a_m)$ and take $\sigma = (a_1a_2)(a_3a_5)$ (using the fact that $n \geq 6$). It follows in both cases that $\tau' = \sigma\tau\sigma^{-1} \neq \tau$, but $\tau'(a_1) = \tau(a_1)$, as desired. □

4.13. Free groups and presentations

We begin with a general definition of a free group by its "universal property".

DEFINITION 4.13.1. A group F is *free* on a subset X if, whenever $f \colon X \to G$ is a function, where G is a group, there exists a unique homomorphism

$$\phi_f \colon F \to G$$

such that $\phi_f(x) = f(x)$ for all $x \in G$. The existence of this unique homomorphism is referred to as the *universal property* of F.

PROPOSITION 4.13.2. *Let F be free on a set X and F' be free on a set X', and suppose $f \colon X \to X'$ is a bijection. Then the homomorphism $\phi_f \colon F \to F'$ given by the universal property is an isomorphism.*

PROOF. Let g be the inverse to f, and let $\phi'_g \colon F' \to F$ be the homorphism given by the universal property for F'. Then $\phi'_g \circ \phi_f(x) = x$ for all $x \in X$ and $\phi_f \circ \phi'_g(x') = x'$ for all $x' \in X$. Since the identity homomorphisms of F and F' also take elements of X and X' to themselves, respectively, the the universal property for F and for F' imply that $\phi'_g \circ \phi_f = \mathrm{id}_F$ and $\phi_f \circ \phi'_g = \mathrm{id}_{F'}$, respectively. Therefore, we have that $\phi'_g = \phi_f^{-1}$, so ϕ_f is an isomorphism. □

EXAMPLE 4.13.3. The integers $\mathbb{Z}$ are a free group on the subset $\{1\}$, since for any group G and element $x \in G$, we can define $\phi \colon \mathbb{Z} \to G$ with $\phi(1) = x$ by $\phi(n) = x^n$ for all $n \in \mathbb{Z}$, and this is the unique homomorphism taking 1 to x.

EXAMPLE 4.13.4. The group $\mathbb{Z}^n$, although a free abelian group, is not a free group. For example, take $n = 2$. Then the map $f \colon \{(1,0),(0,1)\} \to D_3$ with $f(1,0) = r$ and $f(0,1) = s$ cannot be extended to a homomorphism $\phi \colon \mathbb{Z}^2 \to D_3$, for such a function would have to satisfy

$$rs = \phi(0,1)\phi(1,0) = \phi(1,1) = \phi(1,0)\phi(0,1) = sr,$$

which does not hold in D_3.

To show the existence of free groups on larger sets, we construct them explicitly.

DEFINITION 4.13.5. A *word* on in a set X is a symbol

$$x_1^{n_1}x_2^{n_2}\cdots x_k^{n_k}$$

with $x_1, x_2, \dots, x_k \in X$ and $n_1, n_2, \dots, n_k \in \mathbb{Z}$, where $k \geq 0$. If $k = 0$, we sometimes denote this word by e, and it is called the empty word.

REMARK 4.13.6. We write the word x^1 for $x \in X$ more simply as x.

DEFINITION 4.13.7. The *product of two words* $w = x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k}$ and $v = y_1^{m_1} y_2^{m_2} \cdots y_l^{m_l}$ in X is the concatenation

$$w \cdot v = wv = x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k} y_1^{m_1} y_2^{m_2} \cdots y_l^{m_l}.$$

Clearly, concatenation is an associative binary operation on the set of words, and e is an identity element for this operation.

Recall from Example 1.2.25 that there is a smallest equivalence relation containing any relation on a set. So, let us define an equivalence relation on the set of words by a set of generators.

DEFINITION 4.13.8. The standard equivalence relation $\sim$ on the set W_X of words on X is the smallest equivalence relation such that

$$wv \sim wx^0 v \tag{4.13.1}$$

and

$$wx^{m+n} v \sim wx^m x^n v \tag{4.13.2}$$

for all $w, v \in W_X$, $x \in X$, and $m, n \in \mathbb{Z}$.

Two words are then equivalent if and only if one can be obtained from the other by a finite sequence of operations on the word consisting each of adding or removing an x^0 for some $x \in X$ or changing x^{m+n} in a word to $x^m x^n$ or changing $x^m x^n$ in a word to x^{m+n} for some $x \in X$ and $m, n \in \mathbb{Z}$.

DEFINITION 4.13.9. We say that a word $x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k}$ in X is *reduced* if $x_i \neq x_{i+1}$ for all $1 \leq i \leq k-1$ and $n_i \neq 0$ for all $1 \leq i \leq k$.

PROPOSITION 4.13.10. *Every word is equivalent to a unique reduced word.*

PROOF. The relation (4.13.2) tells us recursively for any $n \geq 1$ that

$$wx^n v \sim w(x \cdot x \cdots x)v \quad \text{and} \quad wx^{-n} v \sim w(x^{-1} \cdot x^{-1} \cdots x^{-1})v, \tag{4.13.3}$$

for all $w, v \in W_X$ and $x \in X$, with n symbols "x" appearing on the right-hand sides. If we start with a word w in X, we may use (4.13.3) to expand it and (4.13.1) to remove any 0-powers of elements of X, to obtain an equivalent word of the form

$$x_1^{\pm 1} x_2^{\pm 1} \cdots x_k^{\pm 1}.$$

We may use (4.13.1) and (4.13.3) to remove terms of the form $x_i x_{i+1}^{-1}$ or $x_i^{-1} x_{i+1}$ with $x_i = x_{i+1}$, relabeling after each step, until no such terms exist. We may then gather terms by again applying (4.13.2) to obtain a reduced word equivalent to w.

The process we have described does not change a reduced word. Moreover, the operations of adding in or removing an x^0 from a word or changing x^{m+n} to $x^m x^n$ for some $m, n \in \mathbb{Z}$ or vice-versa do not change the result of the process. Therefore, each word is equivalent to a unique reduced word. □

NOTATION 4.13.11. The set of equivalence classes of words on a set X is denoted F_X.

PROPOSITION 4.13.12. *The set F_X is a group under concatenation of words, and it is generated by the set X.*

PROOF. We give only a sketch. First, we must check that if $w \sim w'$ and $v \sim v'$ are two pairs of equivalent words in X, then $wv \sim w'v'$. This follows quickly from the definition of the equivalence relation $\sim$. So, the binary operation is well-defined, associative, and has identity e. Moreover, the inverse of the equivalence class of a word $x_1^{n_1}x_2^{n_2}\cdots x_k^{n_k}$ is the equivalence class of the word

$$x_k^{-n_k}\cdots x_2^{-n_2}x_1^{-n_1}.$$

That X generates F_X is simply Proposition 2.4.3. □

REMARK 4.13.13. It is typical to denote an element of F_X by any word representing it, which means that we will use the symbol "=" instead of "$\sim$" when interpreting these words as elements of F_X.

EXAMPLES 4.13.14.

a. The free group $F_{\{x\}}$ consists exactly of all x^n for $n \in \mathbb{Z}$, and only $x^0 = e$, so $F_{\{x\}} \cong \mathbb{Z}$.

b. The free group $F_{\{x,y\}}$ with $x \neq y$ consists of all words

$$x^{n_1}y^{m_1}x^{n_1}y^{m_2}\cdots x^{n_k}y^{m_k},$$

where we can take $n_i \neq 0$ for $i \geq 2$ and $m_j \neq 0$ for $j < k$. We have, e.g.,

$$x^2y^{-1}x^{-3}y^{-1}\cdot yx^3y^2x^5 = x^2yx^5.$$

LEMMA 4.13.15. *The group F_X is a free group on the set X.*

PROOF. By Definition 4.13.1, we must show that for any group G and function $f\colon X \to G$, the function

$$\phi_f(x_1^{n_1}x_2^{n_2}\cdots x_k^{n_k}) = f(x_1)^{n_1}f(x_2)^{n_2}\cdots f(x_k)^{n_k}$$

is the unique well-defined homomorphism such that $\phi_f(x) = f(x)$ for all $x \in X$. That ϕ_f is a homomorphism is a direct consequence of its definition and the definition of multiplication of words by concatenation, once it is seen to be well-defined. That it is well-defined is a consequence of the fact that the only relations that are imposed on words are those that exist in any group. That is, for words v and w, $x \in X$, and $m,n \in \mathbb{Z}$, we have

$$\phi_f(vx^0w) = \phi_f(v)\phi_f(x^0)\phi_f(w) = \phi_f(v)f(x)^0\phi_f(w) = \phi_f(v)\phi_f(w),$$
$$\phi_f(vx^{m+n}w) = \phi_f(v)f(x)^{m+n}\phi_f(w) = \phi_f(v)f(x)^m f(x)^n\phi_f(w) = \phi_f(vx^mx^nw),$$

so ϕ_f is constant on equivalent words. □

DEFINITION 4.13.16. The group F_X of Proposition 4.13.12 is the *free group* on a set X.

Proposition 4.13.2 then immediately implies the following.

COROLLARY 4.13.17. *If X and Y are sets with the same cardinality, then F_X and F_Y are isomorphic.*

The following then provides an object that is well-defined up to isomorphism.

NOTATION 4.13.18. The free group on a set with n elements is denoted F_n.

The following gives the relationship between free groups and free abelian groups.

PROPOSITION 4.13.19. *The free abelian group on a set X is isomorphic to the abelianization of the free group on X.*

PROOF. Define $\pi\colon F_X \to \bigoplus_{x \in X} \mathbb{Z}$ by $\pi(x) = e_x$, where e_x is the standard basis element of $\bigoplus_{x \in X} \mathbb{Z}$ corresponding to x. This is a surjective homomorphism. As the image of π is abelian, the map π factors through a sujrective homomorphism $\bar{\pi}\colon F_X^{\mathrm{ab}} \to \bigoplus_{x \in X} \mathbb{Z}$, and therefore the maximal abelian quotient of F_X surjects onto $\bigoplus_{x \in X} \mathbb{Z}$. In F_X^{ab}, we may rearrange the terms of the image of $x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k}$ for $x_i \in X$ and $n_i \in \mathbb{Z}$ with $1 \le i \le k$ so that it is the image of a like element with $x_1, \ldots, x_k$ are all distinct. Now, such an element is sent to $\sum_{i=1}^k n_i e_{x_i}$ under $\bar{\pi}$, and so it is 0 if and only if all $n_i = 0$. If follows that $\bar{\pi}$ is an isomorphism. □

THEOREM 4.13.20. *The free group F_n on n elements cannot be generated by fewer than n elements.*

PROOF. If F_n could be generated by $n-1$ elements, then F_n^{ab} could be generated by $n-1$ elements by Proposition 4.13.19. But F_n^{ab} is isomorphic to the free abelian group on n elements, so Theorem 4.4.16 tells us that F_n^{ab} cannot be generated by $n-1$ elements. □

We omit the proof of the following theorem.

THEOREM 4.13.21. *Every subgroup of a free group is also a free group.*

One of the most important uses of free groups is to give presentations of groups. First, we make the following definition, recalling Lemma 1.2.24 to see that it is well-defined.

DEFINITION 4.13.22. The *normal closure* of a subset S of a group G is the smallest normal subgroup of G containing S, equal to the intersection of all normal subgroups of G containing S.

DEFINITION 4.13.23. A *presentation* of a group G on a set S and a subset T of F_S is a surjective homomorphism $F_S \to G$ with kernel equal to the normal closure of the set T. We say that G is presented by the generating set S and the relation set T, and we write $G \cong \langle S \mid T \rangle$.

REMARK 4.13.24. If G is presented by S and T, then the first isomorphism theorem tells us that $G \cong F_S/R$ by an isomorphism sending $s \in S$ to the coset of s in F_S/R, where R is the normal closure of T.

In fact, we have the following stronger result.

PROPOSITION 4.13.25. *Suppose that G and G' are groups with G presented by S and T. Suppose we are given a subset $\{x_s \mid s \in S\}$ of G'. Then there exists a homomorphism $\phi\colon G \to G'$ with $\phi(s) = x_s$ for all $s \in S$ if and only if the unique homomorphism $\Phi\colon F_S \to G'$ with $\Phi(s) = x_s$ for all $s \in S$ satisfies $T \subseteq \ker\Phi$.*

PROOF. The existence and uniqueness of Φ is by the universal property of F_S. If $T \subseteq \ker\Phi$, then the normal closure R of T is contained in $\ker\Phi$ since $\ker\Phi$ is a normal subgroup of $F_{r,s}$ containing T. In this case, the first isomorphism theorem implies that Φ induces a map $\bar{\Phi}\colon F_S/R \to T$ with $\bar{\Phi}(sR) = x_s$ for all $s \in S$. Since $G \cong F_S/R$ by an isomorphism sending s to sR, the composition $\phi\colon G \to G'$ is the desired map. Similarly, if ϕ exists, then we may compose it with the surjection $F_S \xrightarrow{\sim} G$ taking $s \in S$ to s to obtain a map $\Phi\colon F_S \to G'$ with R, and hence T, in its kernel. □

DEFINITION 4.13.26. If G is presented by finite sets $S = \{s_1, s_2, \dots, s_k\}$ and $T = \{r_1, r_2, \dots, r_d\}$, then G is said to be *finitely presented*.

NOTATION 4.13.27. We write

$$G = \langle s_1, s_2, \dots, s_k \mid r_1, r_2, \dots, r_d \rangle.$$

to denote that G has a presentation by sets $S = \{s_1, s_2, \dots, s_k\}$ and $T = \{r_1, r_2, \dots, r_d\}$.

EXAMPLES 4.13.28. We give several examples of presentations:

a. $F_S \cong \langle S \mid \varnothing \rangle$,

b. $\mathbb{Z}^2 \cong \langle a, b \mid aba^{-1}b^{-1} \rangle$,

c. $\mathbb{Z}/n\mathbb{Z} \cong \langle a \mid a^n \rangle$,

d. $D_n \cong \langle r, s \mid r^n, s^2, rsrs \rangle$.

REMARK 4.13.29. One sometimes writes

$$\langle s_1, s_2, \dots, s_k \mid r_1 = r_1', r_2 = r_2', \dots, r_d = r_d' \rangle$$

for a finite presentation

$$\langle s_1, s_2, \dots, s_k \mid r_1^{-1}r_1', r_2^{-1}r_2', \dots, r_d^{-1}r_d' \rangle.$$

EXAMPLE 4.13.30. We have

$$\mathbb{Z}^n \cong \langle x_1, \dots, x_n \mid x_i x_j = x_j x_i \text{ for } 1 \le i < j \le n \rangle.$$

Note that we can start with the presentation, rather than a group, in order to define new groups.

EXAMPLE 4.13.31. The quaternion group Q_8 is the group of order 8 with the presentation

$$Q_8 = \langle i, j \mid i^4 = e, i^2 = j^2, ij = ji^{-1} \rangle.$$

The elements of Q_8 are usually labelled $\{\pm 1, \pm i, \pm j, \pm k\}$, with $k = ij$, $-1 = i^2$, $-i = i^3$, $-j = j^3$, and $-k = ji = k^3$. We remark that $Q_8 \not\cong D_4$.

Sometimes, we just end up with complicated presentations of familiar groups.

EXAMPLE 4.13.32. Consider the group

$$G = \langle x, y \mid x^2 y, x^4 y \rangle.$$

Then $e = (x^2y)^{-1}x^4y = y^{-1}x^2y$, which forces $x^2 = e$, and then $y = x^2y = e$. Since in fact $x^2 = e$ and $y = e$ imply $x^2y = x^4y = e$, the group G also has a presentation

$$G = \langle x, y \mid x^2, y \rangle,$$

and so is just $\langle x \rangle \cong \mathbb{Z}/2\mathbb{Z}$.

REMARK 4.13.33. To determine whether or not two presentations yield isomorphic groups is a very difficult question. So difficult, in fact, that it has been shown that there does not exist a single recursive computer algorithm into which one can input any two finite group presentations that will always output whether or not they are yield isomorphic groups. Moreover, there exist presentations of groups G from which one cannot write a computer algorithm to determine whether or not a word in the generators of G is equal to the identity of G.

DEFINITION 4.13.34. Let G_1 and G_2 be groups. A free product $G_1 * G_2$ of G_1 and G_2 is a group for which there exist homomorphisms $\iota_i\colon G_i \to G_1 * G_2$ for $i \in \{1,2\}$ such that for any group K and group homomorphisms $\phi_i\colon G_i \to K$ with $i \in \{1,2\}$, there exists a unique homomorphism $\Phi\colon G_1 * G_2 \to K$ with $\Phi \circ \iota_i = \phi_i$ for $i \in \{1,2\}$.

PROPOSITION 4.13.35. *Let G_1 and G_2 be groups. Then the free product of G_1 and G_2 exists and is unique up to isomorphism. Moreover, if G_i has a presentation $G_i \cong \langle S_i \mid T_i \rangle$ for each $i \in \{1,2\}$, then the free product is isomorphic to*

$$G_1 * G_2 \cong \langle S_1 \amalg S_2 \mid T_1 \amalg T_2 \rangle,$$

where $\amalg$ denotes the disjoint union.

PROOF. We verify that the group $N = \langle S_1 \amalg S_2 \mid T_1 \amalg T_2 \rangle$ is a free product of G_1 and G_2. We leave the uniqueness of the free product up to isomorphism as an exercise for the reader. Let $\pi_i\colon F_{S_i} \to G_i$ be the surjections defining the presentation of G_i for $i \in \{1,2\}$. Define homomorphisms $\chi_i\colon F_{S_i} \to N$ by letting $\chi_i(s)$ equal the image of s in N for all $s \in S_i$. By definition of N, we have that T_i is contained in the kernel of χ_i, so the first isomorphism theorem provides maps $\iota_i\colon G_i \to N$ such that $\iota_i(\pi_i(s))$ is the image of s in N for any $s \in S_i$.

Now, for $i \in \{1,2\}$, let $\phi_i\colon G_i \to K$ be a homomorphism to some group K. Then we have a unique map $\Psi\colon F_{S_i \amalg S_j} \to K$ determined by $\Psi(s) = \phi_i(\pi_i(s))$ for all $s \in S_i$ for $i \in \{1,2\}$. If $t \in T_i$ for some i, then $\Psi(t) = \phi_i(e) = e$, so the the first isomorphism theorem yields a homomorphism $\Phi\colon N \to K$ such that $\Phi(\iota_i(g)) = \phi_i(g)$ for all $g = \pi_i(s)$ for some $s \in S$ for $i \in \{1,2\}$. However, the elements of $\pi_i(S_i)$ generate G_i, so we have that $\Phi \circ \iota_i = \phi_i$ for each i. Moreover, Φ is unique, as its values on the images of the elements of $S_1 \amalg S_2$ are determined by the latter equalities. □

REMARK 4.13.36. An element of the free product of groups G and H is an equivalence class of words $g_1h_1g_2h_2\cdots g_kh_k$ with $g_i \in G_i$ and $h_i \in H_i$ for $1 \le i \le k$ (under an equivalence relation under which the identity elements of the two groups are each identified with the identity element of the free product and which otherwise only imposes the relations of the original groups within words), with multiplication induced by concatenation.

CHAPTER 5

Advanced ring theory

5.1. Unique factorization domains

In this section, we investigate the role that prime numbers play in the integers in greater generality. Recall that every nonzero integer can be written as plus or minus a product of distinct prime powers, and these prime powers are unique. Note that the units in $\mathbb{Z}$ are ± 1, so we can say that every nonzero integer can be written as a product of prime powers times a unit. In this section, we investigate this property for a larger class of integral domains.

First, we introduce an analogue of prime numbers.

DEFINITION 5.1.1. Let R be an integral domain. A nonunit and nonzero element $p \in R$ is said to be an *irreducible element* if for every $a, b \in R$ with $p = ab$, either a or b is a unit.

DEFINITION 5.1.2. Two elements a and b of a nonzero commutative ring R with unity are said to *associates* if $a = ub$ with $u \in R^\times$.

Of course, the property of being associate is an equivalence relation on an integral domain R. The equivalence class of 0 is $\{0\}$ and that of 1 is $R^\times$. We have the following simple lemma, which tells us that the equivalence class of an irreducible element consists of irreducible elements.

LEMMA 5.1.3. *If R is an integral domain, and $p \in R$ is irreducible, then so is every associate of p.*

PROOF. That is, if $u \in R^\times$ and $up = ab$ for $a, b \in R$, then $p = (u^{-1}a)b$. As p is irreducible, either $u^{-1}a \in R^\times$ or $b \in R^\times$. Finally, if $u^{-1}a \in R^\times$, then $a \in R^\times$. □

EXAMPLES 5.1.4.

a. The irreducible elements of $\mathbb{Z}$ are $\pm p$ for prime numbers p. The elements p and $-p$ are associates.

b. The irreducible elements of $F[x]$, for a field F, are the irreducible polynomials of F, since the units of $F[x]$ are the nonzero constant polynomials. Every nonzero polynomial has a unique associate with leading coefficient equal to 1.

c. In the subring of $\mathbb{C}$ that is

$$\mathbb{Z}[\sqrt{-2}] = \{a + b\sqrt{-2} \mid a, b \in \mathbb{Z}\},$$

a number of prime integers are no longer irreducible. For instance $2 = -(\sqrt{-2})^2$, and $\sqrt{-2}$ is not a unit. Also, $3 = (1+\sqrt{-2})(1-\sqrt{-2})$, and neither $1+\sqrt{-2}$ nor $1-\sqrt{-2}$ is a unit, for if, e.g., $u \in \mathbb{Z}[\sqrt{-2}]$ with $u(1+\sqrt{-2}) = 1$, then $3u = 1-\sqrt{-2}$, which is clearly impossible. On the other hand, it turns out that 5 is irreducible, though we do not prove this now.

DEFINITION 5.1.5. An integral domain R is a *unique factorization domain*, or a *UFD*, if every nonzero, nonunit element $a \in R$ can be written as a product

$$a = p_1 p_2 \cdots p_r$$

with $p_1, p_2, \dots, p_r$ irreducible elements of R for some $r \geq 1$, and moreover, this expression is unique in the sense that if

$$a = q_1 q_2 \cdots q_s$$

with $q_1, q_2, \dots, q_s$ irreducible for some $s \geq 1$, then $s = r$ and there exists a permutation $\sigma \in S_r$ such that $q_{\sigma(i)}$ and p_i are associates for all $1 \leq i \leq r$.

REMARK 5.1.6. If one wants to allow units, one can rephrase Definition 5.1.5 to read that every nonzero element $a \in R$ can be written as $a = up_1 \cdots p_r$ with $u \in R^\times$ and $p_1, p_2, \dots, p_r$ irreducible in R for some $r = 0$ in a unique manner such that any such decomposition of $a = vq_1 \cdots q_s$ has $s = r$ and, after a reordering of the irreducibles, each q_i is an associate of p_i.

EXAMPLE 5.1.7. The ring $\mathbb{Z}$ is a unique factorization domain.

As we shall see later, $F[x]$ for a field F is a unique factorization domain as well.

EXAMPLE 5.1.8. Consider the subring $F[x^2, xy, y^2]$ of $F[x, y]$. It consists exactly of the polynomials in $F[x, y]$ that can be written as polynomials in x^2, xy, and y^2. These latter three elements are irreducible in $F[x^2, xy, y^2]$, but we have

$$x^2 \cdot y^2 = xy \cdot xy,$$

so factorization is not unique.

A more standard example is the following.

EXAMPLE 5.1.9. Consider the subring $\mathbb{Z}[\sqrt{-5}]$ of $\mathbb{C}$. We have

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

The element 2 divides only elements of the form $a + b\sqrt{-5}$ with $a, b \in \mathbb{Z}$ even, so it does not divide $1 + \sqrt{-5}$ or $1 - \sqrt{-5}$. On the other hand, 2 is irreducible since if $a + b\sqrt{-5}$ divides 2, then so does its complex conjugate, and then

$$(a + b\sqrt{-5})(a - b\sqrt{-5}) = a^2 + 5b^2$$

divides 2, which happens only if $a = \pm 1$ and $b = 0$. Therefore, $\mathbb{Z}[\sqrt{-5}]$ is not a unique factorization domain.

One advantage of unique factorization domains is that they allow us to define a concept of greatest common divisor.

DEFINITION 5.1.10. Let R be a UFD. Let $a_1, a_2, \dots, a_r \in R$ be nonzero. A principal ideal (d) for $d \in R$ is said to be the *greatest common divisor*, or GCD, of $a_1, a_2, \dots, a_r$ if d divides a_i for each $1 \leq i \leq r$ and if d' also divides each a_i, then d' divides d.

The element d in the definition of GCD, if it exists, is only defined up to unit. On the other hand, (d) is independent of this choice.

LEMMA 5.1.11. *Let R be a UFD. Then every collection $a_1, a_2, \dots, a_r$ of nonzero elements of R has a GCD.*

PROOF. We sketch the proof. Factor each a_i into a unit times a product of irreducibles. If there exists an irreducible element p_1 that divides each a_i, an associate of it is one of the irreducibles appearing in the factorization of a_i. We then have $b_i \in R$ with $a_i = pb_i$ for each i, and the factorization of b_i has one fewer irreducible element than that of a_i. We repeat this process until the collection no longer has a common irreducible divisors, obtaining irreducibles $p_1, p_2, \dots, p_k$ such that $d = p_1 p_2 \cdots p_k$ divides every a_i.

We claim that (d) is the GCD of $a_1, a_2, \dots, a_k$. If not, then there exists d' that does not divide d which divides every a_i. This means that there exists an irreducible element $q \in R$ and some $n \geq 1$ such that q^n divides d' but not d. Then q^n divides every a_i, which means since q^n does not divide d that q actually divides each c_i such that $a_i = dc_i$, in contradiction to the definition of d. □

One advantage of having the notion of a GCD is that in quotient fields, it allows us to talk about fractions being in lowest terms.

DEFINITION 5.1.12. Let R be a UFD, and let $a, b \in R$ with $b \neq 0$. We say that the fraction $\frac{a}{b}$ is reduced, or in lowest terms, if the GCD of a and b is (1).

LEMMA 5.1.13. *Let R be a UFD. Every fraction in $Q(R)$ may be written in lowest terms.*

PROOF. Let $a, b \in R$ with $b \neq 0$. Let (d) be the GCD of a and b. Then there exist $a', b' \in R$ with $a = da'$ and $b = db'$, and we have that the GCD of a' and b' is (1). We therefore have that $\frac{a'}{b'} = \frac{a}{b}$, and the former form of the fraction is in lowest terms. □

Let us study factorization in principal ideal domains.

DEFINITION 5.1.14. Let X be a set, and let $\leq$ be a partial ordering on X.

a. An *ascending chain* in X is a sequence $(a_i)_{i \geq 1}$ of elements of X such that $a_i \leq a_{i+1}$ for all $i \geq 1$.

b. We say that X satisfies the *ascending chain condition*, or *ACC*, if every ascending chain $(a_i)_{i \geq 1}$ in X is eventually constant: i.e., there exists $j \geq 1$ such that $a_i = a_j$ for all $i \geq j$.

The following is an equivalent characterization of the ACC.

PROPOSITION 5.1.15. *A nonempty set X with a partial ordering $\leq$ satisfies the ACC if and only if every subset of X contains a maximal element.*

PROOF. If every subset of X contains a maximal element, then clearly ascending chains are eventually constant: i.e., their underlying sets are finite. For the other direction, it suffices to show that if X satisfies the ACC, then it contains a maximal element. Let C be a nonempty chain in X, and suppose it does not have an upper bound. For each $x \in C$, there exists $y \in C$ with $y > x$, as otherwise x would be an upper bound. We may therefore recursively pick $a_i \in X$ with $a_i < a_{i+1}$ for each i, but this is impossible. Thus C has an upper bound, and therefore X has a maximal element by Zorn's lemma. □

DEFINITION 5.1.16. We say that a commutative ring R is *noetherian* if the set of its ideals satisfies the ascending chain condition with respect to containment of ideals.

REMARK 5.1.17. We may rephrase the condition that R be noetherian by saying that if $(I_n)_{n\geq 1}$ is an ascending chain of ideals, then there exists $m \geq 1$ such that the union I of the I_n with $n \geq 1$ equals I_i for all $i \geq m$.

REMARK 5.1.18. One may define a noncommutative ring to be left noetherian (resp., right noetherian rings) if it satisfies the ACC on left ideals (resp., right ideals). In general, a noetherian ring is taken to be one that is both left and right noetherian.

THEOREM 5.1.19. *A commutative ring R is noetherian if and only if every ideal of R is finitely generated.*

PROOF. Suppose that every ideal of R is finitely generated. Let $(I_n)_{n\geq 1}$ be a chain of ideals of R. Let I be the union of the I_n for $n \geq 1$, which is an ideal by Lemma 3.11.10. Since I is finitely generated, $I = (a_1, a_2, \dots, a_r)$, with $a_k \in I$ with $1 \leq k \leq r$ for some $r \geq 1$. For each k, there exists $m_k \geq 1$ with $a_k \in I_{m_k}$, and if we let m be the maximum of the m_k, then $a_k \in I_m$ for every a_k. Since I is the smallest ideal of R containing each a_k, we have $I \subseteq I_m$, which forces $I = I_m$.

Conversely, suppose R is noetherian, and let I be an ideal of R. Let $x_1 \in I$, and suppose inductively that we have constructed $x_1, x_2, \dots, x_n \in R$ with the property that if we set $I_k = (x_1, x_2, \dots, x_k)$ for every $1 \leq k \leq n$, then $I_k \subseteq I_{k+1}$ for every $1 \leq k \leq n-1$. If $I_n \neq I$, then let $x_{n+1} \in I$ with $x_{n+1} \notin I_n$. Then $I_{n+1} = (x_1, x_2, \dots, x_{n+1})$ properly contains I_n. If this process repeats indefinitely, then we have constructed an ascending chain $(I_n)_{n\geq 1}$ that is not eventually constant, which would contradict the assumption that R is noetherian. Therefore, there exists $m \geq 1$ such that $I_m = I$, and so $I = (a_1, a_2, \dots, a_m)$ is finitely generated. □

COROLLARY 5.1.20. *Every principal ideal domain is noetherian.*

PROPOSITION 5.1.21. *Let R be a principal ideal domain. Then every nonzero, nonunit $a \in R$ may be written as $a = p_1 p_2 \cdots p_r$ with the $p_i \in R$ irreducible for all $1 \leq i \leq r$ and some $r \geq 1$.*

PROOF. We claim first that every nonunit $a \in R$ is divisible by an irreducible element of R. If a is not irreducible, set $a_0 = a$ and write $a = a_1 b_1$ with $a_1, b_1 \notin R^\times$. Suppose that a_i divides a_{i-1} for some $i \geq 1$, which implies recursively that a_i divides a. If a_i is irreducible, then we have the claim. If not, then write $a_i = a_{i+1} b_{i+1}$ for some nonunits $a_{i+1}, b_{i+1} \in R^\times$. Since a_{i+1} properly divides a_i, we have that $(a_i) \subsetneq (a_{i+1})$. By Corollary 5.1.20, this process must terminate, which is to say that some a_m is eventually irreducible, and therefore a is divisible by an irreducible element.

Next, we construct another sequence out of our reducible element a. That is, we write $a = a_1 b_1$ with a_1 irreducible, and assume inductively that we have written

$$a = a_1 a_2 \dots a_n b_n$$

with $a_1, a_2, \dots, a_n \in R$ irreducible and nonunit $b_n \in R$ for some $n \geq 0$. If b_n is irreducible for any n, we are done. Otherwise, we obtain a sequence of elements $(b_i)_{i\geq 1}$ with $b_i = a_{i+1} b_{i+1}$ for all $i \geq 1$, which means that $(b_i) \subsetneq (b_{i+1})$ for each i. Again this would contradict the fact that R is noetherian, so eventually the process does terminate, and we have written a as a product of irreducible elements. □

LEMMA 5.1.22. *Let R be a PID, and let $a \in R$ be nonzero. Then (a) is maximal if and only if a is an irreducible element.*

PROOF. Clearly, a cannot be a unit for either condition to hold. If $a = bc$ with b and c non-units, then $(a) \subsetneq (b) \subsetneq R$, so (a) is not maximal. And if (a) is not maximal, then there exists an proper ideal $I = (c)$ of R properly containing (a), so we may write $a = bc$ with $b \in R$. Since the containment is proper, b is not a unit, and c is not a unit by definition. Therefore, a is reducible. □

In a principal ideal domain, irreducible elements play the role that prime numbers play in $\mathbb{Z}$.

LEMMA 5.1.23. *Let R be a PID, and let $p \in R$ be irreducible. If $a, b \in R$ are such that $p \mid ab$, then $p \mid a$ or $p \mid b$.*

PROOF. Let $a, b \in R$ with $p \mid ab$. Then $ab \in (p)$, and (p) is maximal by Lemma 5.1.22. Since every maximal ideal of R is prime, we have that (p) is prime, and therefore either $a \in (p)$ or $b \in (p)$. □

We now prove a key theorem.

THEOREM 5.1.24. *Every principal ideal domain is a unique factorization domain.*

PROOF. Let $a \in R$ be a nonzero, nonunit element. By Proposition 5.1.21, we may write

$$a = p_1 p_2 \cdots p_r$$

with $p_1, p_2, \ldots, p_r$ irreducible. We have only to show that this decomposition is unique in the appropriate sense. So, suppose that

$$a = q_1 q_2 \cdots q_s$$

with $q_1, q_2, \ldots, q_s$ irreducible. If $r = 1$, then a is irreducible, so $s = 1$ and $p_1 = q_1$. Suppose by induction we have proven uniqueness whenever there is a decomposition of a with fewer than $r \geq 2$ irreducibles. In particular, we may assume that $s \geq r$.

As a consequence of Lemma 5.1.23, we have that p_r divides some q_i for some $1 \leq i \leq s$. Since q_i is irreducible, this means that $q_i = wp_r$ with $w \in R^\times$. Since R is an integral domain, we then have

$$p_1 p_2 \cdots p_{r-1} = wq_1 q_2 \cdots q_{i-1} q_{i+1} \cdots q_s.$$

As $s \geq 2$ by assumption, note that wq_1 is an associate to q_1 and the expression on the right is a product of $s-1$ irreducible elements. By induction, we have $r = s$, and there exists a bijective function

$$\sigma\colon \{1, 2, \ldots, r-1\} \to \{1, 2, \ldots, i-1, i+1, \ldots, r\}$$

with $q_{\sigma(i)}$ and p_i associates for each $1 \leq i \leq r-1$. We may extend σ to an element of S_r by setting $\sigma(r) = i$, and then $q_{\sigma(r)} = q_i$ is an associate of p_r as well, proving uniqueness. □

Given that every polynomial ring over a field is a PID, we have the following corollary. It is an interesting exercise to prove it directly.

COROLLARY 5.1.25. *For any field F, the ring $F[x]$ is a unique factorization domain.*

Corollary 3.10.2 tells us what we may already have known from experience, that we can factor one-variable polynomials into irreducible factors over a field, and there is only one way to do this.

5.2. Polynomial rings over UFDs

Now that we know that every PID is a UFD, the question arises: is every UFD also a PID? The answer, in fact, is no. For this, let us examine polynomial rings over integral domains in a bit more detail.

DEFINITION 5.2.1. Let R be an integral domain. A polynomial $f \in R[x]$ is said to be *primitive* if the only elements of R that divide all of the coefficients of f are units.

In a UFD, we can actually talk about the GCD of the coefficients of a polynomial.

DEFINITION 5.2.2. Let R be a UFD. The *content* of the a polynomial in $R[x]$ is the GCD of its coefficients.

REMARK 5.2.3. If R is a UFD, then a polynomial in $R[x]$ is primitive if and only if the GCD of its coefficients is (1).

DEFINITION 5.2.4. A polynomial in $R[x]$ for a nonzero ring R with unity is said to be *monic* is its leading coefficient is 1.

REMARK 5.2.5. Monic polynomials in $R[x]$, where R is a UFD, are primitive.

LEMMA 5.2.6. *Let R be a UFD. If (c) is the content of $f \in R[x]$ for, then there exists a primitive polynomial $g \in R[x]$ with $f = cg$.*

PROOF. By definition, c divides each coefficient of f, so $f = cg$ for some $g \in R[x]$. Let $d \in R$ be such that (d) is the content of g. Then $g = dh$ for some $h \in R[x]$, so we have $f = cdh$. But this implies that cd divides every coefficient of f, so cd divides the content c, forcing d to be a unit. Therefore, g is primitive. □

EXAMPLE 5.2.7. The polynomial $f = 25x^2 + 10x - 15$ in $\mathbb{Z}[x]$ has content 5, and so it is not primitive. In fact, $f = 5g$, where $g = 5x^2 + 2x - 3$, and g is primitive.

LEMMA 5.2.8 (Gauss's Lemma). *Let R be a UFD. Then the product of any two primitive polynomials in $R[x]$ is primitive.*

PROOF. Let

$$f = \sum_{i=0}^{n} a_i x^i \quad \text{and} \quad g = \sum_{j=0}^{m} b_j x^j$$

be primitive polynomials in $R[x]$. The kth coefficient of fg is $c_k = \sum_{i=0}^{k} a_i b_{k-i}$. If p is an irreducible element of R, then since f and g are primitive, there exist minimal nonnegative integers r and s such that $p \nmid a_r$ and $p \nmid b_s$. Since $p \mid a_i$ for $i < r$ and $p \mid b_j$ for $j < s$, which is to say that $p \mid b_{r+s-i}$ for $r < i \le r+s$, we have that p divides every term of c_{r+s} except $a_r b_s$, which it does not divide. Therefore, p does not divides c_{r+s}. Since p was arbitrary, fg is primitive. □

Note that we can speak about polynomials being irreducible in $R[x]$ for any integral domain R, since we have a notion of irreducible element in such a ring. For a field F, this coincides with the usual notion of an irreducible polynomial.

PROPOSITION 5.2.9. *Let R be an integral domain, and let $F = Q(R)$.*

a. If $f \in R[x]$ is a primitive polynomial that is irreducible as an element of $F[x]$, then f is irreducible in $R[x]$. In particular, if f cannot be written as a product of two nonconstant polynomials in $R[x]$, then it is irreducible in $R[x]$.

b. Suppose that R is a UFD. If $f \in R[x]$ is irreducible, then it is irreducible as an element of $F[x]$ as well. In fact, if $f \in R[x]$ and $f = gh$ for nonconstant $g,h \in F[x]$, then there exists $\alpha \in F^\times$ such that $g' = \alpha g$ and $h' = \alpha^{-1}h$ are in $R[x]$ and therefore $f = g'h'$ in $R[x]$.

PROOF. First, we treat part a. If $f \in R[x]$ is primitive and $f \in R[x]$ is reducible (which is to say, not irreducible and not a unit or zero), then we can write $f = gh$ for nonunits $g,h \in R[x]$. If g or h is constant, then f is not primitive, so neither is constant, and therefore f is reducible in $F[x]$.

Next, we turn to part b. Suppose that $f \in R[x]$ can be written as $f = gh$ with $g,h \in F[x]$ nonconstant. Let (d) (resp., (e)) be a multiple of all of the denominators of the coefficients of g (resp., h), written in lowest terms. Then $def = g'h'$, where $g',h' \in R[x]$ are nonconstant. The content of def is contained in (de), so the content of $g'h'$ is as well. By unique factorization in R, we may write $de = d'e'$, where $d' \in R$ divides the content of g' and e' divides the content of h', and we may then divide g' by d' and h' by e' to obtain g'' and h'' in $R[x]$ such that $f = g''h''$. Therefore, f is reducible in $R[x]$, and the remaining statement of the lemma holds as well. □

We are now ready to prove the following.

THEOREM 5.2.10. *If R is a UFD, then $R[x]$ is a UFD as well.*

PROOF. Let $f \in R[x]$ be a nonzero element that is not a unit. Write

$$f = f_1 f_2 \dots f_r$$

with $f_i \in R[x]$ nonconstant, where r is maximal such that this can be done. Note that such a maximal r exists as the degree of f is finite. For $1 \le i \le r$, let (c_i) be the content of f_i, and define $g_i \in R[x]$ by $f_i = c_i g_i$. Set $c = c_1 c_2 \cdots c_r$, and set $g = g_1 g_2 \cdots g_r$. Now, if any g_i were not irreducible in $F[x]$ for $F = Q(R)$, then it would not be irreducible in $R[x]$ by Proposition 5.2.9b. Moreover, since g_i is primitive, it would then be written as a product of two nonconstant polynomials in $R[x]$, which would contradict the maximality of r. Therefore, each g_i is irreducible. Since R is a UFD, we may also write $c = p_1 p_2 \cdots p_k$ with $p_i \in R$ irreducible for $1 \le i \le k$ and some $k \ge 0$, and so

$$f = p_1 p_2 \cdots p_k g_1 g_2 \cdots g_r$$

is a factorization of f into irreducibles in $R[x]$.

Now, if

$$f = q_1 q_2 \cdots q_l h_1 h_2 \cdots h_s$$

with $q_i \in R$ irreducible and $h_i \in R[x]$ irreducible and nonconstant, then $(q_1 q_2 \cdots q_l)$ is the content of f by Gauss's lemma, and so $q_1 q_2 \cdots q_l$ agrees with c up to unit in R. Since R is a UFD, it follows that $l = k$ and there exists $\sigma \in S_k$ such that each $q_{\sigma(i)}$ is an associate of p_i. Next, we have

$$g_1 g_2 \cdots g_r = u h_1 h_2 \cdots h_s$$

for some unit $u \in R^\times$, and by uniqueness of factorization in $F[x]$, we have that $s = r$, and there exists $\tau \in S_r$ such that $h_{\tau(i)} = v_i g_i$ for some $v_i \in F^\times$ for each $1 \leq i \leq r$. But the content of each g_i and each h_j is (1), since these elements are irreducible in $R[x]$, and therefore writing $v_i = \frac{a_i}{b_i}$ with $a_i, b_i \in R$, the fact that $b_i h_{\tau(i)} = a_i g_i$ implies that $(a_i) = (b_i)$, since both sides must have the same content. In other words, $v_i \in R^\times$, and so $h_{\tau(i)}$ and g_i are associates in $R[x]$, finishing the proof of uniqueness. □

EXAMPLES 5.2.11.

a. Since $\mathbb{Z}$ is a UFD, so is $\mathbb{Z}[x]$. However, $\mathbb{Z}[x]$ is not a PID, since (p,x) is not principal.

b. Since $\mathbb{Q}[x]$ is a UFD, so is $\mathbb{Q}[x,y]$. Again, $\mathbb{Q}[x,y]$ is not a PID, since (x,y) is not principal.

c. If R is any UFD, then $R[x_1, x_2, \cdots x_n]$ is a UFD for any $n \geq 1$.

5.3. Irreducibility of polynomials

In this section, we investigate criteria for determining if a polynomial is irreducible or not.

DEFINITION 5.3.1. Let R be an integral domain. We say that a polynomial $f = \sum_{i=0}^n a_i x^i$ be a polynomial in $R[x]$ that satisfies $a_n \notin \mathfrak{p}$, $a_i \in \mathfrak{p}$ for all $0 \leq i \leq n-1$, and $a_0 \notin \mathfrak{p}^2$ for some $n \geq 1$ and prime ideal $\mathfrak{p}$ in R is an Eisenstein polynomial (with respect to $\mathfrak{p}$).

THEOREM 5.3.2 (Eistenstein criterion). *Let R be an integral domain, and let $f \in R[x]$ be an Eiseinstein polynomial.*

a. If R is a UFD, then f is irreducible in $Q(R)[x]$.

b. If f is primitive, then it is irreducible in $R[x]$.

PROOF. Suppose $f = \sum_{i=0}^n a_i x^i$ is of degree n and Eisenstein with respect to a prime ideal $\mathfrak{p}$ of R. By Proposition 5.2.9, it suffices for each part to show that f is not a product of two nonconstant polynomials in $R[x]$. So, let $g = \sum_{i=0}^s b_i x^i$ and $h = \sum_{j=0}^t c_j x^j$ be polynomials in $R[x]$ with $f = gh$, where $s + t = n$. We then have

$$a_k = \sum_{i=0}^k b_i c_{k-i}$$

for all $0 \leq k \leq n$. In particular, $a_0 = b_0 c_0$ is an element of $\mathfrak{p}$ but not $\mathfrak{p}^2$. Since $\mathfrak{p}$ is prime, at least one of b_0 and c_0 lies in $\mathfrak{p}$, but as $a_0 \notin \mathfrak{p}^2$, at least one does not lie in $\mathfrak{p}$ as well.

Without loss of generality, suppose that $b_0 \in \mathfrak{p}$ and $c_0 \notin \mathfrak{p}$. As $a_n = b_s c_t \notin \mathfrak{p}$, we have $b_s \notin \mathfrak{p}$. Let $k \geq 1$ be minimal such that $b_k \notin \mathfrak{p}$. If $k < n$, then $a_k \in \mathfrak{p}$ and $b_i \in \mathfrak{p}$ for $i < k$, so we have $b_k c_0 \in \mathfrak{p}$, which therefore forces $c_0 \in \mathfrak{p}$ by the primality of $\mathfrak{p}$. Therefore, $k = n$, which means that h is constant, proving the result. □

We will most commonly be concerned with the Eisenstein criterion in the case that $R = \mathbb{Z}$.

EXAMPLE 5.3.3. For any prime number p and integer $n \geq 1$, the polynomial $x^n - p$ is irreducible by the Eisenstein criterion. That is, we take our prime ideal to be (p) in the ring $\mathbb{Z}$.

EXAMPLE 5.3.4. For a prime number p, set

$$\Phi_p = \frac{x^p - 1}{x-1} = x^{p-1} + x^{p-2} + \cdots + 1.$$

This polynomial has as its roots in $\mathbb{C}$ the distinct pth roots of unity that are not equal to 1. Over $\mathbb{Q}$, we claim it is irreducible. For this, consider the polynomial

$$\Phi_p(x+1) = \frac{(x+1)^p - 1}{x} = \sum_{i=0}^{p-1} \binom{p}{i+1} x^i,$$

which has coefficents divisible by p but not p^2 except for its leading coefficient a_{p-1}, which is 1. Therefore, $\Phi_p(x+1)$ is Eisenstein, hence irreducible. But if Φ_p were to factor into g and h, then $\Phi_p(x+1)$ would factor into $g(x+1)$ and $h(x+1)$, which have the same leading coefficients as g and h, and hence are nonconstant if and only if g and h are. In other words, Φ_p is irreducible as well.

REMARK 5.3.5. The condition in the Eisenstein criterion that the constant coefficient not lie in the square of the prime ideal is in general necessary. For instance, $x^2 - p^2 \in \mathbb{Z}[x]$ is never irreducible for a prime p.

Often, we can tell if a polynomial is irreducible by considering its reductions modulo ideals.

PROPOSITION 5.3.6. *Let R be an integral domain, and let $\mathfrak{p}$ be a prime ideal of R. Let $f \in R[x]$ with leading coefficient not in $\mathfrak{p}$. Let $\bar{f}$ denote the image of f in $(R/\mathfrak{p})[x]$ given by reducing its coefficients modulo $\mathfrak{p}$.*

a. *If R is a UFD and $\bar{f}$ is irreducible in $Q(R/\mathfrak{p})[x]$, then f is irreducible in $Q(R)[x]$.*

b. *If f is primitive and $\bar{f}$ is irreducible in $R/\mathfrak{p}[x]$, then f is irreducible in $R[x]$.*

PROOF. If R is a UFD and f is reducible in $Q(R)[x]$, then by Proposition 5.2.9, we have that $f = gh$ for some nonconstant $g,h \in R[x]$. Similarly, if f is primitive and reducible in $R[x]$, then $f = gh$ for nonconstant $g,h \in R[x]$. In either case, since the leading coefficient of f is not in $\mathfrak{p}$ and $\mathfrak{p}$ is prime, we have that the leading coefficients of g and h are not in $\mathfrak{p}$ as well. That is, the images of g and h in $(R/\mathfrak{p})[x]$ are nonconstant, which means that $\bar{f}$ is a product of two nonconstant polynomials, hence reducible in $Q(R/\mathfrak{p})[x]$. □

REMARK 5.3.7. For $R = \mathbb{Z}$, Proposition 5.3.6 tells us in particular that if $f \in \mathbb{Z}[x]$ is monic and its reduction $\bar{f} \in \mathbb{F}_p[x]$ modulo p is irreducible for any prime p, then f is irreducible.

EXAMPLE 5.3.8. Let $f = x^4 + x^3 + 1001 \in \mathbb{Z}[x]$. We claim that f is irreducible in $\mathbb{Q}[x]$. For this, consider its reduction modulo 2. The polynomial $\bar{f} = x^4 + x^3 + 1 \in (\mathbb{Z}/2\mathbb{Z})[x]$ is either irreducible, has a root in $(\mathbb{Z}/2\mathbb{Z})[x]$, or is a product of two irreducible polynomials of degree 2. But $\bar{f}(0) = \bar{f}(1) = 1$, and $x^2 + x + 1$ is the only irreducible polynomial of degree 2 in $(\mathbb{Z}/2\mathbb{Z})[x]$, and $(x^2+x+1)^2 = x^4+x^2+1 \neq \bar{f}$, so $\bar{f}$ is irreducible. By Proposition 5.3.6, f is irreducible in $\mathbb{Q}[x]$.

EXAMPLE 5.3.9. The converse to Proposition 5.3.6 does not hold. For instance, x^2+x+1 is irreducible in $\mathbb{Q}[x]$, but it has a root in $(\mathbb{Z}/3\mathbb{Z})[x]$.

We also have the following simple test for the existence of roots of polynomials over UFDs.

PROPOSITION 5.3.10. *Let R be a UFD and $f = \sum_{i=0}^{n} a_i x^i \in R[x]$ with $a_0, a_n \neq 0$. Suppose that $\alpha \in Q(R)$ is a root of f, and write α in reduced form as $\alpha = \frac{c}{d}$ for some $c,d \in R$. Then c divides a_0 and d divides a_n in R.*

PROOF. Since $x - \frac{c}{d}$ divides f in $Q(R)[x]$ and $\frac{c}{d}$ is in reduced form, it follows from Proposition 5.2.9 that $f = (dx-c)g$ for some $g \in R[x]$. Writing $g = \sum_{i=0}^{n-1} b_i g_i$, we see that $a_0 = -cb_0$ and $a_n = db_{n-1}$. □

EXAMPLE 5.3.11. Let $f = 2x^3 - 3x + 5 \in \mathbb{Z}[x]$. We check that $f(1) = 4$, $f(-1) = 6$, $f(5) \equiv -10 \bmod 25$, $f(-5) \equiv 20 \bmod 25$, and $f(\frac{1}{2})$, $f(-\frac{1}{2})$, $f(\frac{5}{2})$, and $f(-\frac{5}{2})$ are all represented by reduced fractions with denominators equal to 4. Proposition 5.3.10 therefore tells us that f has no roots in $\mathbb{Q}$, hence is irreducible, being of degree 3.

5.4. Euclidean domains

DEFINITION 5.4.1. A *norm* f on an ring R is a function $f\colon R \to \mathbb{Z}_{\geq 0}$ with $f(0) = 0$. We say that f is *positive* if the only $a \in R$ for which $f(a) = 0$ is $a = 0$.

DEFINITION 5.4.2. Let R be an integral domain. A *Euclidean norm* ν on R is a norm on R such that for all nonzero $a, b \in R$, one has

i. $\nu(a) \leq \nu(ab)$, and

ii. there exist $q, r \in R$ with $a = qb + r$ and either $\nu(r) < \nu(b)$ or $r = 0$.

REMARK 5.4.3. Property (ii) of Definition 5.4.2 is known as the division algorithm.

DEFINITION 5.4.4. A *Euclidean domain* R is an integral domain such that there exists a Euclidean norm on R.

EXAMPLES 5.4.5.

a. The integers $\mathbb{Z}$ are a Euclidean domain with Euclidean norm $\nu(a) = |a|$ for any nonzero $a \in \mathbb{Z}$.

b. Every polynomial ring $F[x]$ over a field F is a Euclidean domain, the degree function providing a Euclidean norm on $F[x]$.

LEMMA 5.4.6. *In a Euclidean domain R with Euclidean norm ν, the minimal value of ν on all nonzero elements of R is $\nu(1)$, and $\nu(u) = \nu(1)$ for $u \in R$ if and only if $u \in R^{\times}$.*

PROOF. By the definition of a Euclidean norm, we have $\nu(1) \leq \nu(a \cdot 1) = \nu(a)$ for all nonzero $a \in R$. If $u \in R^{\times}$, then $\nu(u) \leq \nu(u \cdot u^{-1}) = \nu(1)$, so $\nu(u) = \nu(1)$. Conversely, if $b \in R$ with $\nu(b) = \nu(1)$, then we may write $1 = qb + r$ for some $q, r \in R$ with either $\nu(r) < \nu(1)$ or $r = 0$. By what we have shown, the latter holds, so $qb = 1$, and b is a unit. □

EXAMPLE 5.4.7. In $F[x]$, the units are exactly the nonzero constant polynomials, i.e., those with degree 0.

While we will explain below that not every PID is a Euclidean domain, it is the case that every Euclidean domain is a PID.

THEOREM 5.4.8. *Every Euclidean domain is a PID.*

PROOF. Let I be a nonzero ideal in a Euclidean domain R with Euclidean norm ν. We must show that I is principal. Let $b \in I$ be a nonzero element with minimal norm among all elements of I. For any $a \in I$, we may write $a = qb + r$ with $q, r \in R$ and either $\nu(r) < \nu(b)$ or $r = 0$. Note that $a, b \in I$, so $r \in I$ as well, which precludes the possibility of $\nu(r) < \nu(b)$, since $\nu(r)$ is minimal among norms of elements of I. Therefore, we have $r = 0$, so $a \in (b)$. As a was arbitrary and $b \in I$, we have $I = (b)$. □

The key property of Euclidean domains is the ability to perform the Euclidean algorithm, which we see in the following.

THEOREM 5.4.9 (Euclidean algorithm). *Let R be a Euclidean domain with Euclidean norm ν, and let $a, b \in R$ be nonzero elements. Let $r_{-1} = a$ and $r_0 = b$. Suppose recursively that we are given elements $r_j \in R$ for $-1 \leq j \leq i$ and some $i \geq 0$. If $r_i \neq 0$, write*

$$r_{i-1} = q_{i+1}r_i + r_{i+1} \tag{5.4.1}$$

with $q_{i+1}, r_{i+1} \in R$ and either $\nu(r_{i+1}) < \nu(r_i)$ or $r_{i+1} = 0$. If $r_{i+1} \neq 0$, repeat the process with i replaced by $i+1$. The process terminates with $d = r_n \neq 0$ and $r_{n+1} = 0$ for some $n \geq 1$, and (d) is the GCD of a and b. Moreover, we may use the formulas in (5.4.1) *and recursion to write d as $d = xa + yb$ for some $x, y \in R$.*

PROOF. We note that the process must terminate, as the values of the $\nu(r_i)$ for $i \geq 0$ are decreasing. Moreover, the result $d = r_n$ satisfies $r_{n-1} = q_{n+1}r_n$, so it divides r_{n-1} by definition, and then we see by downward recursion using (5.4.1) that d divides every r_{i-1}. Finally, if c is any common divisor of a and b, then it again recursively divides each r_i (this time by upwards recursion and (5.4.1)), so c divides d. Therefore, (d) is the GCD of a and b.

Note that $d = r_{n-2} - q_n r_{n-1}$, and suppose that we may write $d = zr_j + wr_{j+1}$ for some $-1 \leq j \leq n-2$. If $j = -1$, we are done. Otherwise, note that $r_{j+1} = r_{j-1} - q_{j+1}r_j$, so

$$d = zr_j + w(r_{j-1} - q_{j+1}r_j) = wr_{j-1} + (z - q_{j+1}w)r_j,$$

and we have written d as an R-linear combination of r_{j-1} and r_j. Repeat the process for $j-1$. The final result is the desired R-linear combintation of a and b. □

EXAMPLE 5.4.10. Take $\mathbb{Z}$ and its usual Euclidean norm. We take $a = 550$ and $b = 154$. Then $550 = 3 \cdot 154 + 88$, so we set $r_1 = 88$. Then $154 = 88 + 66$, so we set $r_2 = 66$, and $88 = 66 + 22$, so we set $r_3 = 22$, and $66 = 3 \cdot 22$, so we stop at $d = r_3 = 22$, which is therefore the greatest common divisor of a and b. Working backwards, we obtain

$$22 = 88 - 66 = 88 - (154 - 88) = 2 \cdot 88 - 154 = 2 \cdot (550 - 3 \cdot 154) - 154 = 2 \cdot 550 - 7 \cdot 154.$$

That is, we have written d as $a + (-4)b$.

Often Euclidean norms come in the form of multiplicative norms.

DEFINITION 5.4.11. A *multiplicative norm* $N\colon R \to \mathbb{Z}_{\geq 0}$ on a commutative ring R with unity is a positive norm such that for all $N(ab) = N(a)N(b)$ for all $a, b \in R$.

REMARK 5.4.12. Note that the existence of a multiplicative norm N on a commutative ring R with unity forces R to be an integral domain, for if $ab = 0$, then $N(a)N(b) = N(ab) = 0$, so either $N(a) = 0$ or $N(b) = 0$, and therefore either $a = 0$ or $b = 0$.

EXAMPLE 5.4.13. The absolute value on $\mathbb{Z}$ is a multiplicative norm, as well as a Euclidean norm.

EXAMPLE 5.4.14. The function N on the Gaussian integers $\mathbb{Z}[i]$ given by $N(a+bi) = a^2+b^2$ is a multiplicative norm. Clearly, $a^2+b^2 = 0$ if and only if $a+bi = 0$. Given $a,b,c,d \in \mathbb{Z}$, we have

$$\begin{aligned} N((a+bi)(c+di)) = (ac-bd)^2 + (ad+bc)^2 &= (ac)^2 + (bd)^2 + (ad)^2 + (bc)^2 \\ &= (a^2+b^2)(c^2+d^2) = N(a+bi)N(c+di). \end{aligned}$$

PROPOSITION 5.4.15. *The ring $\mathbb{Z}[i]$ of Gaussian integers is a Euclidean domain with respect to the Euclidean norm $N(a+bi) = a^2+b^2$ for $a, b \in \mathbb{Z}$.*

PROOF. Since N is a multiplicative norm, we need only check the division algorithm. Extend N to a function on $\mathbb{C}$ by defining $N(a+bi) = a^2+b^2$ for $a,b \in \mathbb{R}$. Let $a,b,c,d \in \mathbb{Z}$ with $(c,d) \neq (0,0)$. Then we have

$$\frac{a+bi}{c+di} = s+ti$$

for some $s,t \in \mathbb{Q}$, and let $e, f \in \mathbb{Z}$ be integers with $|s-e| \leq 1/2$ and $|t-f| \leq 1/2$. Then we have

$$\begin{aligned} N(a+bi-(e+fi)(c+di)) &= N(c+di)N((s-e)+(t-f)i) \\ &\leq N(c+di)\left(\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2\right) = N(c+di)/2 < N(c+di), \end{aligned}$$

so the division algorithm is satisfied: $a+bi = q(c+di)+r$ with $q = e+fi$ and $N(r) < N(c+di)$ if $r \neq 0$. □

COROLLARY 5.4.16. *The units in $\mathbb{Z}[i]$ are exactly* $1, -1, i, -i$.

PROOF. Since N is a Euclidean norm on $\mathbb{Z}[i]$, the units are exactly those nonzero elements of norm $N(1) = 1$. We have $a^2+b^2 = 1$ if and only if $(a,b) = (\pm 1, 0)$ or $(a,b) = (0, \pm 1)$. □

LEMMA 5.4.17. *If $a,b,c,d \in \mathbb{Z}$ and $c+di$ divides $a+bi$ in $\mathbb{Z}[i]$, then $c-di$ divides $a-bi$ in $\mathbb{Z}[i]$.*

PROOF. Write $a+bi = (c+di)(e+fi)$ for some $e, f \in \mathbb{Z}$. Then $a = ce-df$ and $b = cd+de$, so

$$(c-di)(e-fi) = (ce-df) - (cf+de)i = a-bi.$$

□

We can completely determine the irreducible elements in $\mathbb{Z}[i]$ as follows.

PROPOSITION 5.4.18. *The irreducible elements in $\mathbb{Z}[i]$ are, up to multiplication by a unit, $1+i$, primes $p \in \mathbb{Z}$ with $p \equiv 3 \bmod 4$, and $a+bi$ for $a,b \in \mathbb{Z}$ such that $p = a^2+b^2 \equiv 1 \bmod 4$ is a prime in $\mathbb{Z}$. Moreover, the primes in $\mathbb{Z}$ that can be written in the form a^2+b^2 are exactly 2 and those that are 1 modulo 4.*

PROOF. First, note that if $a+bi$ divides $c+di$ in $\mathbb{Z}[i]$ for integers a,b,c,d, then $N(a+bi)$ divides $N(c+di)$, since N is multiplicative. So, $1+i$ is irreducible since $N(1+i) = 2$.

Let p be an odd prime in $\mathbb{Z}$. If p is divisible by some irreducible element $\pi = a+bi$ with $a,b \in \mathbb{Z}$, then since p is prime, only one of two things can happen. Either $ab = 0$, or a and b are relatively prime in $\mathbb{Z}$, noting Corollary 5.4.16. Suppose $ab \neq 0$. By Lemma 5.4.17, we have that $a-bi$ divides p, and $\bar{\pi} = a-bi$ is irreducible. If $\bar{\pi}$ were associate to π, then π would divide $2a = (a+bi)+(a-bi)$ and $2b = -i((a+bi)-(a-bi))$. Then π divides 2, but that is impossible. Thus, π and $\bar{\pi}$ both dividing p implies that p is divisible by $N(\pi) = a^2+b^2$. As p is prime, we have $p = a^2+b^2$.

So, we have shown that either our odd prime p is irreducible in $\mathbb{Z}[i]$ or $p = a^2+b^2$ for some $a,b \in \mathbb{Z}$. Note that the squares in $\mathbb{Z}/4\mathbb{Z}$ are 0 and 1, so any integer of the form a^2+b^2 is 0, 1, or 2 modulo 4. In particular, if $p \equiv 3 \bmod 4$, then p is irreducible in $\mathbb{Z}[i]$.

If $p \equiv 1 \bmod 4$ is prime in $\mathbb{Z}$, then $(\mathbb{Z}/p\mathbb{Z})^\times$ has order divisible by 4. As $\mathbb{Z}/p\mathbb{Z}$ contains only two roots of x^2-1, which are -1 and 1, so $(\mathbb{Z}/p\mathbb{Z})^\times$ contains an element of order 4. In particular, there exists $n \in \mathbb{Z}$ such that $n^2 \equiv -1 \bmod p$, which is to say that p divides n^2+1. If p were irreducible in $\mathbb{Z}[i]$, then p would divide either $n+i$ or $n-i$, but then it would divide both, being an integer. Thus p would divide $2i$, which it does not. So, p is reducible, which means equals a^2+b^2 for some $a,b \in \mathbb{Z}$. □

LEMMA 5.4.19. *Let N be a multiplicative norm on an integral domain R. Then $N(u) = 1$ for all $u \in R^\times$.*

PROOF. We have $N(1) = N(1)^2$, and R is an integral domain, so $N(1) = 1$. Moreover, since

$$N(u^{-1})N(u) = N(1) = 1,$$

we have that $N(u^{-1}) = N(u)^{-1}$, and therefore $N(u) = 1$. □

EXAMPLE 5.4.20. Consider the multiplicative norm N on $\mathbb{Z}[\sqrt{-5}]$ given by

$$N(a+b\sqrt{-5}) = |a^2+5b^2|.$$

We have $a^2+5b^2 = 1$ if and only if $a = \pm 1$ and $b = 0$, so the only units in $\mathbb{Z}[\sqrt{-5}]$ are ± 1. Now, if $2 = \alpha\beta$ for some nonunits $\alpha,\beta \in \mathbb{Z}[\sqrt{-5}]$, then $4 = N(2) = N(\alpha)N(\beta)$, so $N(\alpha) = 2$, but 2 is clearly not a value of N. Therefore, 2 is irreducible, and so is 3. Also, we have that $N(1 \pm \sqrt{-5}) = 6$, and since 2 and 3 are not values of N, we have that $1 \pm \sqrt{-5}$ is irreducible as well. As these elements are all non-associates, the existence of the two factorizations

$$6 = 2 \cdot 3 = (1+\sqrt{-5})(1-\sqrt{-5})$$

proves that $\mathbb{Z}[\sqrt{-5}]$ is not a UFD.

Not all principal ideal domains are Euclidean. We give most of the outline of how one produces an example.

DEFINITION 5.4.21. An nonzero, non-unit element b of an integral domain R is called a *universal side divisor* if every element $a \in R$ may be written in the form $a = qb + r$ for some $q, r \in R$ with $r = 0$ or $r \in R^\times$.

LEMMA 5.4.22. *Let R be a Euclidean domain with Euclidean norm ν. Let $b \in R$ be a nonzero, non-unit element such that $\nu(b)$ is minimal among nonzero, non-unit elements of R. Then b is a universal side divisor of R.*

PROOF. Let $a \in R$. By definition of ν, we may write $a = qb + r$ with $\nu(r) < \nu(b)$ or $r = 0$. By the minimality of $\nu(b)$, we must have that r is a unit or 0. □

EXAMPLE 5.4.23. We claim that the ring $R = \mathbb{Z}[(1+\sqrt{-19})/2]$ is not Euclidean. Suppose by contradiction that it is a Euclidean domain, and let ν be a Euclidean norm on R. We also have the multiplicative norm N on R given by

$$N\left(a + b\frac{1+\sqrt{-19}}{2}\right) = \left(a + b\frac{1+\sqrt{-19}}{2}\right)\left(a + b\frac{1-\sqrt{-19}}{2}\right) = a^2 + ab + 5b^2. \tag{5.4.2}$$

Note that if $\alpha \in R - \mathbb{Z}$, then $N(\alpha) \geq 5$, so the only units in R are ± 1.

Let $\beta \in R$ be a universal side divisor, which exists as R is Euclidean, and write $2 = q\beta + r$ for $q \in R$ and $r \in \{0, 1, -1\}$. We then have that $N(\beta)$ divides $N(2-r)$ as N is multiplicative, so $N(\beta)$ divides 4 or 9, and this implies $\beta \in \{\pm 2, \pm 3\}$ by the formula for N. Now take $\alpha = (1 + \sqrt{-19})/2$, and set $\alpha = q'\beta + r'$ with $q' \in R$ and $r' \in \{0, 1, -1\}$. We have $N(\alpha) = N(\alpha - 1) = 5$ and $N(\alpha + 1) = 7$, which are not multiples of $N(\beta) \in \{4, 9\}$, so we obtain a contradiction.

DEFINITION 5.4.24. A *Dedekind-Hasse norm* on an integral domain R is a positive norm μ on R such that for every $a, b \in R$, either $a \in (b)$ or there exists a nonzero element $c \in (a, b)$ such that $\mu(c) < \mu(b)$.

PROPOSITION 5.4.25. *An integral domain R is a PID if and only if there exists a Dedekind-Hasse norm on R.*

PROOF. Suppose first that μ is a Dedekind-Hasse norm on R. Let I be a nonzero ideal of R, and let $b \in I - \{0\}$ with minimal norm under μ. If $a \in I$, then since there does not exist a nonzero element $c \in (a, b) \subseteq I$ with $\mu(c) < \mu(b)$ by the minimality of $\mu(b)$, we have by definition of a Dedekind-Hasse norm that $a \in (b)$. Thus $I = (b)$.

Suppose on the other hand the R is a PID. Define $\mu\colon R \to \mathbb{Z}_{\geq 0}$ by $\mu(0) = 0$, $\mu(u) = 1$ for $u \in R^\times$, and $\mu(p_1 p_2 \cdots p_k) = 2^k$ if $p_1, \ldots, p_k$ are irreducible elements of R. This is well-defined as R is a UFD. Given $a, b \in R$, we have $(a, b) = (d)$ for some $d \in R$, since R is a PID. Since d divides b, we have $\mu(d) \leq \mu(b)$. If $\mu(d) = \mu(b)$, then a and b have the same number of divisors as d and therefore are associates, so $a \in (b)$. Thus, μ is a Dedekind-Hasse norm. □

EXAMPLE 5.4.26. We have already seen that $R = \mathbb{Z}[(1+\sqrt{-19})/2]$ is not a Euclidean domain. To see that R is a PID, it suffices to show that the multiplicative norm N on R given by (5.4.2) is a Dedekind-Hasse norm on R. We outline the standard unenlightening verification.

Let $\alpha,\beta \in R$ with $\alpha \notin (\beta)$. We claim that there exist $s,t \in R$ with $0 < N(s\alpha - t\beta) < N(\beta)$. Note that we can extend N to a map $N\colon Q(R) \to \mathbb{Z}_{\geq 0}$ by the formula (5.4.2), allowing $a,b \in \mathbb{Q}$. Our condition that N on R be a Dedekind-Hasse norm is then that $0 < N(s\frac{\alpha}{\beta} - t) < 1$. We will find s and t. For this, write

$$\frac{\alpha}{\beta} = \frac{a+b\sqrt{-19}}{c}$$

for $a,b,c \in \mathbb{Z}$ with no common divisor and $c > 1$.

First one considers the cases with $c \geq 4$. If $c = 2$, then either a or b is odd, then take $s = 1$ and $t = ((a-1)+b\sqrt{-19})/2$. If $c = 3$, then $a^2 + 19b^2 \not\equiv 0 \bmod 3$, so $a^2 + 19b^2 = 3q + r$ with $r \in \{1,2\}$. Take $s = a - b\sqrt{-19}$ and $t = q$. If $c = 4$, then again either a or b is odd. If only one is, then write $a^2 + 19b^2 = 4q + r$ with $1 \leq r \leq 4$, and take $s = a - b\sqrt{-19}$ and $t = q$. If both are, write $a^2 + 19b^2 = 8q + 4$, and take $s = \frac{1}{2}(a - b\sqrt{-19})$ and $t = q$.

Now suppose that $c \geq 5$. Since $(a,b,c) = (1)$, we have $x,y,z \in \mathbb{Z}$ such that $xa + yb + zc = 1$. Write $ay - 19bx = qc + r$, with $q \in \mathbb{Z}$ and $|r| \leq c/2$. Take $s = y + x\sqrt{-19}$ and $t = q - z\sqrt{-19}$. The reader will check that

$$N\left(s\frac{\alpha}{\beta} - t\right) = c^{-2}N\left(s(a+b\sqrt{-19}) - tc\right) = \frac{r^2+19}{c^2},$$

which is at most $\frac{1}{4} + \frac{19}{36} = \frac{7}{9}$ if $c \geq 6$ and at most $\frac{4}{25} + \frac{19}{25} = \frac{23}{25}$ if $c = 5$.

5.5. Vector spaces over fields

In this section, we give a very brief discussion of the theory of vector spaces over fields, as it shall be subsumed by the sections that follow it.

DEFINITION 5.5.1. Let F be a field. A *vector space* V over F is an abelian group under addition that is endowed with an operation $\cdot\colon F \times V \to V$ of scalar multiplication such that for all $a,b \in F$ and $v,w \in V$, one has

i. $1 \cdot v = v$,

ii. $a \cdot (b \cdot v) = (ab) \cdot v$,

iii. $(a+b) \cdot v = a \cdot v + b \cdot v$,

iv. $a \cdot (v+w) = a \cdot v + a \cdot w$.

REMARK 5.5.2. In a vector space V over a field F, we typically write av for $a \cdot v$, where $a \in F$ and $v \in V$.

EXAMPLE 5.5.3. If F is a field, then F^n is a vector space over F under the operation

$$a \cdot (\alpha_1, \alpha_2, \ldots, \alpha_n) = (a\alpha_1, a\alpha_2, \ldots, a\alpha_n)$$

for $a, \alpha_1, \alpha_2, \ldots, \alpha_n \in F$.

DEFINITION 5.5.4. An element of a vector space V over a field F is called a *vector*, and the elements of F under in the operation $\cdot$ are referred to as *scalars*.

EXAMPLE 5.5.5. In every vector space V, there is an element 0, and it is called the zero vector.

DEFINITION 5.5.6. The *zero vector space* 0 is the vector space over any field F that is the set $\{0\}$ with the operation $a \cdot 0 = 0$ for all $a \in F$.

EXAMPLE 5.5.7. If F is a field, then $F[x]$ is a vector space over F with $a \cdot f$ for $a \in F$ and $f \in F[x]$ defined to be the usual product of polynomials in $F[x]$. I.e., the operation of scalar multiplication is just multiplication by a constant polynomial.

EXAMPLE 5.5.8. The field $\mathbb{C}$ is an $\mathbb{R}$-vector space, as well as a $\mathbb{Q}$-vector space. The field $\mathbb{R}$ is a $\mathbb{Q}$-vector space. The operations of scalar multiplication are just restrictions of the usual multiplication map on $\mathbb{C}$.

The reader will easily check the following.

LEMMA 5.5.9. *If V is a vector space over a field F, then for $a \in F$ and $v \in V$, we have*

a. $0 \cdot v = 0$,

b. $a \cdot 0 = 0$,

c. $-(av) = (-a)v = a(-v)$.

DEFINITION 5.5.10. Let V be a vector space over a field F. A *subspace* W of V is a subset that is closed under the operations of addition and scalar multiplication to W (i.e., to maps $W \times W \to V$ and $F \times W \to V$, respectively) and is a vector space with respect to these operations.

The following is easily proven.

LEMMA 5.5.11. *A subset W of a vector space V is a subspace if and only if it is a subgroup under addition and closed under scalar multiplication.*

EXAMPLES 5.5.12.

a. The zero subspace $\{0\}$ and V are both subspaces of any vector space V.

b. The field F is a subspace of $F[x]$.

DEFINITION 5.5.13. Let V be a vector space over a field F, and let S be a subset of V. A *linear combination* of elements of S is any sum

$$\sum_{i=1}^{n} a_i v_i$$

with $v_1, v_2, \ldots, v_n$ distinct vectors in S and $a_1, a_2, \ldots, a_n \in F$ for some $n \geq 0$. We say that such a linear combination is *nontrivial* if there exists a j with $1 \leq j \leq n$ and $a_j \neq 0$.

DEFINITION 5.5.14. Let V be a vector space over a field F and S be a set of vectors in V. The subspace *spanned* by S, also known as the *span* of V, is the set of all linear combinations of elements of S, or simply the zero subspace if S is empty.

EXAMPLE 5.5.15. For any vector space V, the set V spans V.

DEFINITION 5.5.16. We say that a set S of vectors in a vector space V over a field F *spans* V if V equals the subspace spanned by S.

That is, S spans an F-vector space V if, for every $v \in V$, there exist $n \geq 0$, $v_i \in V$, and $a_i \in F$ for $1 \leq i \leq n$ such that

$$v = \sum_{i=1}^{n} a_i v_i.$$

DEFINITION 5.5.17. We say that a set of S of vectors in a vector space V over a field F is *linearly independent* if every nontrivial linear combination of vectors in S is nonzero. Otherwise, S is said to be *linearly dependent*.

That is, a set S of vectors in an F-vector space V is linearly independent if whenever $n \geq 1$, $v_i \in V$ and $a_i \in F$ for $1 \leq i \leq n$ and

$$\sum_{i=1}^{n} a_i v_i = 0,$$

then $a_i = 0$ for all $1 \leq i \leq n$.

LEMMA 5.5.18. *Let S be a linearly independent subset of a vector space V over a field F, and let W be the span of F. If $v_0 \in V - W$, then $S \cup \{v_0\}$ is also linearly independent.*

PROOF. Let $v_1, v_2, \ldots, v_n \in S$ and $c_0, c_1, \ldots, c_n \in F$ for some $n \geq 1$, and suppose that

$$\sum_{i=0}^{n} c_i v_i = 0.$$

We cannot have $c_0 \neq 0$, as then

$$v_0 = -c_0^{-1} \sum_{i=1}^{n} c_i v_i \in W.$$

On the other hand, the fact that $c_0 = 0$ implies that $c_i = 0$ for all $1 \leq i \leq n$ by the linear independence of V. Thus, $S \cap \{v_0\}$ is linearly independent. □

EXAMPLE 5.5.19. In any vector space V, the empty set is linearly independent. If $v \in V$ is nonzero, then $\{v\}$ is also a linearly independent set.

DEFINITION 5.5.20. A subset B of a vector space V over a field F is said to be a *basis* of V over F if it is linearly independent and spans V.

EXAMPLE 5.5.21. The set $\{e_1, e_2, \ldots, e_n\}$ of F^n, where e_i is the element of F^n that has a 1 in its ith coordinate and 0 in all others, is a basis of F^n.

EXAMPLE 5.5.22. The set $\{x^i \mid i \geq 0\}$ is a basis of $F[x]$. That is, every polynomial can be written as a finite sum of distinct monomials in a unique way.

REMARK 5.5.23. For a field F, it is very hard to write down a basis of $\prod_{i=0}^{\infty} F$. In fact, the proof that it has a basis uses the axiom of choice.

DEFINITION 5.5.24. A vector space V is said to be *finite dimensional* if it has a finite basis (i.e., a basis with finitely many elements). Otherwise V is said to be *infinite dimensional*.

The following theorem employs Zorn's lemma.

THEOREM 5.5.25. *Let V be a vector space over a field F. Every linearly independent subset of V is contained in a basis of V.*

PROOF. Let S be a linearly independent subset of V, and let X denote the set of linearly independent subsets of V that contain S. We order X by containment of subsets. If $\mathscr{C}$ is a chain in X, then its union $U = \bigcup_{T \in \mathscr{C}} T$ is linearly independent since if $v_1, v_2, \ldots, v_n \in U$ for some $n \geq 1$, then each v_i is contained in some $T_i \in X$ for each $1 \leq i \leq n$, and one of the sets T_j contains the others, since $\mathscr{C}$ is a chain. Since T_j is linearly independent, any nontrivial linear combination of the elements v_i with $1 \leq i \leq n$ is nonzero. Therefore, U is linearly independent as well, so is contained in X.

By Zorn's Lemma, X now contains a maximal element B, and we want to show that B spans V, so is a basis of V containing S. Let W denote the span of B. If $v_0 \in V - W$, then $B' = B \cup \{v_0\}$ is linearly independent by Lemma 5.5.18, so an element of X, which contradicts the maximality of B. That is, $V = W$, which is to say that B spans V. □

In particular, the empty set is contained in a basis of any vector space, so we have the following:

COROLLARY 5.5.26. *Every vector space over a field contains a basis.*

A similar argument yields the following.

THEOREM 5.5.27. *Let V be a vector space over a field F. Every subset of V that spans V contains a basis of V.*

PROOF. Let S be a spanning subset of V. Let X denote the set of linearly independent subsets of S, and order X by containment. As seen in the proof of Theorem 5.5.25, any union of a chain of linearly independent subsets is linearly independent, so has an upper bound. Thus, Zorn's lemma tells us that X contains a maximal element B. Again, we want to show that B spans V, so is a basis. If it were not, then there would exist some element of V which is not in the span of B, but is in the span of S. In particular, there exists an element $v_0 \in S$ that is not in the span of B. The set $B \cup \{v_0\}$ is linearly independent, contradicting the maximality of B. □

We also have the following, which can be generalized to a statement on cardinality.

THEOREM 5.5.28. *Let V be a vector space over a field F. If V is finite dimensional, then every basis of V contains the same number of elements, and otherwise every basis of V is infinite.*

PROOF. Let $B_1 = \{v_1, v_2, \ldots, v_n\}$ be a basis of V with a minimal number n of elements, and let $B = \{w_1, w_2, \ldots, w_m\}$ be another basis of V with $m \geq n$. Then B_1 spans V, so w_1 is a nontrivial linear combination of the v_i for $1 \leq i \leq n$:

$$w_1 = \sum_{i=1}^{n} a_i v_i \tag{5.5.1}$$

for some $a_i \in F$. Letting j be such that $a_j \neq 0$, we may write v_j as a linear combination of w_1 and the v_i with $i \neq j$. In other words, $B_2 = (B_1 - \{v_j\}) \cup \{w_1\}$ spans V. Suppose

$$c_j w_1 + \sum_{\substack{i=1 \\ i \neq j}}^{n} c_i v_i = 0 \tag{5.5.2}$$

for some $c_i \in F$. Using (5.5.1), we may rewrite the sum in (5.5.2) as a linear combination of the v_i, the coefficient of v_j in which is $a_j c_j$, which forces $c_j = 0$ as B_1 is a linearly independent set. But then we see from (5.5.2) that all $c_i = 0$ as $B - \{v_j\}$ is linearly independent. So, B_2 is a basis of V.

Suppose by recursion that, for $k \leq m$, we have found a basis B_k of order n of V that contains only $w_1, \ldots, w_{k-1}$ and elements of B. Then w_k is a nontrivial linear combination of the elements of B_k, and the coefficient of some v_l is nonzero in this linear combination by the linear independence of B. We therefore have that $B_{k+1} = (B_k - \{v_l\}) \cup \{w_k\}$ spans V, and a similar argument to the above shows that it is a basis. Finally, we remark that the basis B_{m+1} must be B_1 itself, since it contains B_1, so we have $m = n$, as desired. □

DEFINITION 5.5.29. The *dimension* of a finite-dimensional vector space V over a field F is the number of elements in a basis of V over F. We write $\dim_F(V)$ for this dimension.

EXAMPLE 5.5.30. The space F^n is of dimension n over F.

The maps between vector spaces that respect the natural operations on the spaces are called linear transformations.

DEFINITION 5.5.31. A *linear transformation* $T\colon V \to W$ of F-vector spaces is a function from V to W satisfying

$$T(v+v') = T(v) + T(v') \quad \text{and} \quad T(av) = aT(v)$$

for all $a \in F$ and $v, v' \in V$

REMARK 5.5.32. In other words, a linear transformation is a homomorphism of the underlying groups that "respects scalar multiplication."

DEFINITION 5.5.33. A linear transformation $T\colon V \to W$ of F-vector spaces is an isomorphism of F-vector spaces if it is there exists an linear transformation $T^{-1}\colon W \to V$ that is inverse to it.

Much as with group and ring homomorphisms, we have the following:

LEMMA 5.5.34. *A linear transformation is an isomorphism if and only if it is a bijection.*

EXAMPLES 5.5.35. Let V and W be F-vector spaces.

a. The identity map $\mathrm{id}_V\colon V \to V$ is an F-linear transformation (in fact, isomorphism).

b. The zero map $0\colon V \to W$ is an F-linear transformation.

5.6. Modules over rings

DEFINITION 5.6.1. Let R be a ring. A *left R-module*, or *left module over R*, is an abelian group M together with an operation $\cdot\colon R \times M \to M$ such that for all $a, b \in R$ and $m, n \in M$, one has

i. $1 \cdot m = m$,

ii. $(a \cdot b) \cdot m = (ab) \cdot m$,

iii. $(a+b) \cdot m = a \cdot m + b \cdot n$,

iv. $a \cdot (m+n) = a \cdot m + a \cdot n$.

DEFINITION 5.6.2. Let R be a commutative ring. We refer more simply to a *left R-module* as a *R-module*, or *module over R*.

REMARK 5.6.3. When one speaks simply of a module over a ring R, one means by default a left R-module.

NOTATION 5.6.4. When an abelian group M is seen as a left module over a ring R via the extra data of some operation $R \times M \to M$, we say that this operation endows M with the additional structure of a left R-module.

EXAMPLE 5.6.5. The definition of a module over a field coincides with the definition of a vector space over a field. In other words, to say that M a module over a field F is exactly to say that M is a vector space over F.

EXAMPLE 5.6.6. The modules over $\mathbb{Z}$ are exactly the abelian groups. That is, suppose that A is a $\mathbb{Z}$-module, which by definition is an abelian group with an additional operation $\cdot\colon \mathbb{Z} \times A \to A$. We show that this additional operation satisfies $n \cdot a = na$ for $n \in \mathbb{Z}$ and $a \in A$, where na is the usual element of the abelian group A. So, let $a \in A$. By axiom (i), we have $1 \cdot a = a$, and then the distributivity of axiom (iii) allows us to see that $n \cdot a = na$ for all $n \geq 1$. Using axioms (iv) and (ii), we have

$$0 \cdot a = 0 \cdot (2a - a) = 0 \cdot 2a - 0 \cdot a = (0 \cdot 2) \cdot a - 0 \cdot a = 0 \cdot a - 0 \cdot a = 0,$$

and then finally we have

$$(-n) \cdot a + n \cdot a = (n-n) \cdot a = 0 \cdot a = 0,$$

so $(-n) \cdot a = -na$ for $n \geq 1$.

EXAMPLE 5.6.7. For a ring R and $n \geq 1$, the direct product R^n is a left $M_n(R)$-module via matrix multiplication $(A, v) \mapsto A \cdot v$ for $A \in M_n(R)$ and $v \in R^n$,viewing elements of R^n as column vectors.

We also have the notion of a right R-module.

DEFINITION 5.6.8. Let R be a ring. A *right R-module*, or *right module over R*, is an abelian group M together with an operation $\cdot\colon M \times R \to R$ such that for all $a, b \in R$ and $m, n \in M$, one has

i. $m \cdot 1 = m$,

ii. $m \cdot (a \cdot b) = m \cdot (ab)$,

iii. $m \cdot (a + b) = m \cdot a + n \cdot b$,

iv. $(m + n) \cdot a = m \cdot a + n \cdot a$.

EXAMPLE 5.6.9. Every left ideal I over a ring R is a left R-module with respect to the restriction $R \times I \to I$ of the multiplication on R. Every right ideal over R is a right module with respect to the restriction $I \times R \to I$ of the multiplication on R.

DEFINITION 5.6.10. Let R be a ring. The opposite ring R^{op} to R is the ring that is the abelian group R together with the multiplication $\cdot^{\mathrm{op}} \colon R \times R \to R$ given by $a \cdot^{\mathrm{op}} b = ba$, where the latter product is taken in R.

REMARK 5.6.11. The identity map induces an isomorphism $R \to (R^{\mathrm{op}})^{\mathrm{op}}$ of rings.

The reader will easily check the following.

LEMMA 5.6.12. *A right module M over R also has the structure of a left module over R^{op}, where the latter operation $\cdot^{\mathrm{op}} \colon R^{\mathrm{op}} \colon M \to M$ is given by $a \cdot^{\mathrm{op}} m = ma$, where the latter product is that given by the right R-module structure of M.*

EXAMPLE 5.6.13. For a field F, the map $T \colon M_n(F) \to M_n(F)$ given by transpose (that is, $A \mapsto A^T$ for $A \in M_n(F)$) is a ring isomorphism between $M_n(F)$ and $M_n(F)^{\mathrm{op}}$.

We also have the notion of a bimodule.

DEFINITION 5.6.14. Let R and S be rings. An abelian group M that is a left R-module and a right S-module is called an *R-S-bimodule* if

$$(r \cdot m) \cdot s = r \cdot (m \cdot s)$$

for all $r \in R$, $s \in S$, and $m \in M$.

EXAMPLES 5.6.15.

a. Any left R-module M over a commutative ring R is an R-R-bimodule with respect to given left operation and the (same) right operation $m \cdot r = rm$ for $m \in M$ and $r \in R$.

b. A two-sided ideal of a ring R is an R-R-bimodule with respect to the operations given by the usual multiplication on R.

c. For $m, n \geq 1$, the abelian group $M_{mn}(R)$ of m-by-n matrices with entries in R is an $M_m(R)$-$M_n(R)$-bimodule for the operations of matrix multiplication.

Let us return our focus to R-modules, focusing on the case of left modules, as right modules are just left modules over the opposite ring by Lemma 5.6.12.

DEFINITION 5.6.16. An *R-submodule* (or, *submodule*) N of a left module M over a ring R is a subset of N that is closed under addition and the operation of left R-multiplication and is an R-module with respect to their restrictions $+ \colon N \times N \to N$ and $\cdot \colon R \times N \to N$ to N.

LEMMA 5.6.17. *Let R be a ring, M be a left R-module, and N be a subset of M. Then N is an R-submodule of M if and only if it is nonempty, closed under addition, and closed under left R-multiplication.*

PROOF. Clearly, it suffices to check that if N is nonempty and closed under addition and left R-multiplication, then it is an R-submodule. The condition of being closed under left R-multiplication assures that 0 and inverses of elements of N lies in N, so N is an abelian group under $+$ on M. The axioms for N to be an R-module under $\cdot$ are clearly satisfied as they are satisfied by elements of the larger set M. □

EXAMPLES 5.6.18.

a. The subspaces of a vector space V over a field F are exactly the F-submodules of V.

b. The subgroups of an abelian group are the $\mathbb{Z}$-submodules of that group.

c. Any left ideal I of R is a left R-submodule of R viewed as a left R-module.

d. Any intersection of R-submodules is an R-submodule as well.

e. For an R-module M and a left ideal I, the abelian group

$$IM = \left\{ \sum_{i=1}^{n} a_i m_i \mid a_i \in I, m_i \in M \text{ for } 1 \le i \le n \right\}$$

is an R-submodule of M.

We also have the following construction.

DEFINITION 5.6.19. Let M be an R-module and $\{N_i \mid i \in I\}$ be a collection of submodules for an indexing set I. The *sum* of the submodules N_i is the submodule $\sum_{i \in I} N_i$ of M with elements $\sum_{i \in I} n_i$ for $n_i \in N_i$ and all but finitely many n_i equal to 0.

If M is an R-module and N is a submodule, we may speak of the quotient abelian group M/N. It is an R-module under the action $r \cdot (m+N) = rm+N$ for $r \in R$ and $m \in M$. This is well-defined, as a different representative $m+n$ of the coset $m+N$ for $n \in N$ will satisfy $r(m+n)+N = rm+rn+N = rm+N$.

DEFINITION 5.6.20. Let M be a left R-module and N be an R-submodule of M. The *quotient module* M/N of M by N is the abelian group of cosets together with the multiplication $R \times M/N \to M/N$ given by $r \cdot (nN) = (rn)N$.

EXAMPLE 5.6.21. For an R-module M and a left ideal I, we have the quotient module M/IM. In particular, note that R/I is a left R-module with respect to $r(s+I) = rs+I$, even if it is not a ring (i.e., if I is not two-sided).

We can also speak of homomorphisms of R-modules.

DEFINITION 5.6.22. Let M and N be left modules over a ring R. A left R-module homomorphism $\phi \colon M \to N$ is a function such that $\phi(r \cdot m) = r\phi(m)$ and $\phi(m+n) = \phi(m) + \phi(n)$ for all $r \in R$ and $m, n \in M$.

NOTATION 5.6.23. If R is commutative (or it is understood that we are working with left modules), we omit the word "left" and speak simply of R-module homomorphisms.

REMARK 5.6.24. A right R-module homomorphism $\phi \colon M \to N$ is just a left R^{op}-module homomorphism.

DEFINITION 5.6.25. Let M and N be left modules over a ring R.

a. An *isomorphism* $f\colon M \to N$ of left R-modules is a bijective homomorphism.

b. An *endomorphism* of a left R-module M is a homomorphism $f\colon M \to M$ of left R-modules.

c. An *automorphism* of a left R-modules M is an isomorphism $f\colon M \to M$ of left R-modules.

NOTATION 5.6.26. Sometimes, we refer to an R-module homomorphism as an R-linear map, and an endomorphism of R-modules as an R-linear endomorphism.

EXAMPLES 5.6.27.

a. The zero map $0\colon M \to M$ and the identity map $\mathrm{id}\colon M \to M$ are endomorphisms of an R-module M, with id being an automorphism.

b. Let V and W be vector spaces over a field F. A left F-module homomorphism $\phi\colon V \to W$ is just an F-linear transformation.

c. Let N be an R-submodule of a left R-module M. The inclusion map $\iota_N\colon N \to M$ is an R-module homomorphism, as is the quotient map $\pi_N\colon M \to M/N$.

d. If M is an R-S-bimodule, then right multiplication $\psi_s\colon M \to M$ by an element $s \in S$ defines a left R-module endomorphism. In particular, if R is a commutative ring, then multiplication by $r \in R$ defines an R-module endomorphism. Note that if R is noncommutative, then the condition that left multiplication by $r \in R$ be a left module homomorphism $M \to M$ is that $r(sm) = s(rm)$ for all $r, s \in R$ and $m \in M$, which need not hold.

e. The identity map $F^n \to F^n$ provides an isomorphism between F^n viewed as a left $M_n(F)$-module via $(A, v) \mapsto Av$ for $A \in M_n(F)$ and $v \in F^n$ (viewing v as a column vector) and F^n viewed as a left $M_n(F)^{\mathrm{op}}$-module via $(A, v) \mapsto v^T A$.

Note that we may speak of the kernel and the image of a left R-module, as an R-module homomorphism is in particular a group homomorphism. The reader will easily verify the following.

LEMMA 5.6.28. *Let $\phi\colon M \to N$ be a left R-module homomorphism. Then* $\ker\phi$ *and* $\operatorname{im}\phi$ *are R-submodules of M and N, respectively.*

We also have analogues of all of the isomorphism theorems for groups. Actually, these are virtually immediate consequences of said isomorphism theorems, as the fact that one has isomorphisms of groups follows immediately from them, and then one need only note that these isomorphisms are actually homomorphisms of R-modules.

THEOREM 5.6.29. *Let R be a ring. Let $\phi\colon M \to N$ be an homomorphism of left R-modules. Then there is an isomorphism $\bar{\phi}\colon M/\ker\phi \to \operatorname{im}\phi$ given by $\bar{\phi}(m + \ker\phi) = \phi(m)$.*

THEOREM 5.6.30. *Let R be a ring, and let N and N' be left R-submodules of an R-module N. Then there is an isomorphism of R-modules*

$$M/(M \cap N) \xrightarrow{\sim} (M+N)/N, \qquad m + (M \cap N) \mapsto m + N.$$

THEOREM 5.6.31. *Let R be a ring, let M be an R-module, and let $Q \subseteq N$ be R-submodules of M. Then there is an isomorphism*

$$M/N \xrightarrow{\sim} (M/Q)/(N/Q), \qquad m + N \mapsto (m+Q) + (N/Q).$$

We also have the following analogue of Theorems 2.13.10 and 3.8.23.

THEOREM 5.6.32. *Let R be a ring, let M be an R-modules, and let N be an R-submodule of M. Then the map $P \mapsto P/N$ gives a bijection between submodules P of M containing N and submodules of M/N. This bijection has inverse $Q \mapsto \pi_N^{-1}(Q)$ on submodules Q of M/N, where $\pi_N \colon M \to M/N$ is the quotient map.*

5.7. Free modules and generators

DEFINITION 5.7.1. Let S be a subset of an R-module M.

a. The submodule of M *generated* by S is the intersection of all submodules of M containing S.

b. We say that S *generates* M, or is a *set of generators* or *generating set* of M, if no proper R-submodule of M contains S.

REMARK 5.7.2. The R-submodule of M generated by S consists of the elements $\sum_{i=1}^{n} a_i m_i$ with $m_i \in S$ and $a_i \in R$ for $1 \le i \le n$ and some $n \ge 1$. The proof is much as before.

REMARK 5.7.3. The sum $\sum_{i \in I} N_i$ of submodules N_i of M is the submodule generated by $\cup_{i \in I} N_i$.

NOTATION 5.7.4. The R-submodule of an R-module M generated by for a single element $m \in M$ (or, more precisely, by $\{m\}$) is denoted $R \cdot m$.

DEFINITION 5.7.5. We say that an R-module is *finitely generated* if it has a finite set of generators.

DEFINITION 5.7.6. We say that an R-module is *cyclic* if it can be generated by a single element.

EXAMPLE 5.7.7. A cyclic R-submodule of R is just a principal left ideal.

We can define direct sums and direct products of modules.

DEFINITION 5.7.8. Let $(M_i)_{i \in I}$ be a collection of left modules over a ring R.

a. The *direct product* $\prod_{i \in I} M_i$ is the R-module that is the direct product of the abelian groups M_i together with the left R-multiplication $r \cdot (m_i)_{i \in I} = (rm_i)_{i \in I}$ for $r \in R$ and $m_i \in M_i$ for all $i \in I$.

b. The *direct sum* $\bigoplus_{i \in I} M_i$ is the R-module that is the direct sum of the abelian groups M_i together with the left R-multiplication $r \cdot (m_i)_{i \in I} = (rm_i)_{i \in I}$ for $r \in R$ and $m_i \in M_i$ for all $i \in I$ with all but finitely many $m_i = 0$.

REMARK 5.7.9. If I is a finite set, then the canonical injection

$$\bigoplus_{i \in I} M_i \to \prod_{i \in I} M_i$$

is an isomorphism. In this case, the two concepts are often used interchangeably.

NOTATION 5.7.10. A direct sum (resp., product) of two R-modules M and N is denoted $M \oplus N$.

DEFINITION 5.7.11. We say that an R-submodule A of an R-module B is a *direct summand* of C if there exists an R-module C such that $B = A \oplus C$. In this case, C is called a *complement* to A in B.

DEFINITION 5.7.12. Let R be a ring.

a. An R-module M is *free* on a subset X of M if for any R-module N and function $\bar{\phi} \colon X \to N$ of elements of N, there exists a unique R-module homomorphism $\phi \colon M \to N$ such that $\phi(x) = \bar{\phi}(x)$ for all $x \in X$.

b. A *basis* of an R-module M is a subset of M on which it is free.

REMARK 5.7.13. An abelian group A is free on a set X if and only if it is a free $\mathbb{Z}$-module on X, as follows from Proposition 4.4.11.

In fact, we have the following alternative definition of a free R-module. The proof is nearly identical to Proposition 4.4.11, so omitted.

PROPOSITION 5.7.14. *An R-module M is free on a basis X if and only if the set X generates M and, for every $n \geq 1$ and $x_1, x_2, \dots, x_n \in X$, the equality*

$$\sum_{i=1}^{n} c_i x_i = 0$$

for some $c_1, c_2, \dots, c_n \in R$ implies that $c_i = 0$ for all i.

REMARK 5.7.15. We might refer to the property that a set X generates an R-module M as saying that M is the R-span of X. The property that $\sum_{i=1}^{n} c_i x_i = 0$ implies $c_i = 0$, where $c_i \in R$ and $x_i \in X$ for $1 \leq i \leq n$ and some $n \geq 1$ can be referred to as saying that the set X is R-linearly independent.

COROLLARY 5.7.16. *For any set X, the R-module $\bigoplus_{x \in X} R$ is free on the standard basis $\{e_x \mid x \in X\}$, where e_x for $x \in X$ is the element which is nonzero only in its x-coordinate, in which it is* 1.

PROOF. The e_x span $\bigoplus_{x \in X} R$ by its definition and are clearly R-linearly independent. □

COROLLARY 5.7.17. *Every R-module is a quotient of a free R-module.*

PROOF. Let M be an R-module, and choose a generating set X of M (e.g., M itself). Take the unique R-module homomorphism

$$\psi \colon \bigoplus_{x \in X} R \to M$$

which satisfies $\psi(e_x) = x$ for all $x \in X$. It is onto as X generates M. □

Noting Corollary 5.5.26, we also have the following.

COROLLARY 5.7.18. *Every vector space over a field F is a free F-module.*

The following is also a consequence of the universal property. Though we restrict to the finite case, it can be improved to a statement on cardinality.

THEOREM 5.7.19. *Let R be a commutative ring. A free module M on a set X is isomorphic to a free module N on a set Y if and only if X and Y have the same cardinality.*

PROOF. If X and Y have the same cardinality, then any bijection $f\colon X \to Y$ gives an injection $X \to N$ which extends uniquely to a homomorphism $\phi\colon M \to N$. Similarly, the inverse of f extends uniquely to a homomorphism $\psi\colon N \to M$, and $\psi \circ \phi$ (resp., $\phi \circ \psi$) is then the unique extension to a homomorphism of the inclusion $X \to M$ (resp., $Y \to N$), therefore the identity. That is, ϕ and ψ are inverse isomorphisms.

For the converse, we first suppose that Y is infinite and that there is an isomorphism $M \to N$. Let B denote the image of X in N, which is then necessarily an R-basis of N. Each element $y \in Y$ is contained in the span of a finite subset B_y of B. The union B' of these sets B_y spans Y. For any $v \in B - B'$, the set $B' \cup \{v\}$ is R-linearly dependent, which cannot happen as B is a basis. Thus, $B = B'$. Now, the cardinality $|B|$ of B is at most the cardinality of the disjoint union of the sets B_y for $y \in Y$, each of which is finite. In particular, we have

$$|X| = |B| \le |Y \times \mathbb{Z}| = |Y|,$$

the latter equality holding as Y is infinite. If X is also infinite, then by reversing the roles of X and Y, this forces $|X| = |Y|$.

Finally, suppose that Y is finite, without loss of generality. Let $\mathfrak{m}$ be a maximal ideal of R. Consider the field $F = R/\mathfrak{m}$, and observe that

$$M/\mathfrak{m}M \cong \left(\bigoplus_{x\in X} R\right)/\mathfrak{m}\left(\bigoplus_{x\in X} R\right) \cong \bigoplus_{x\in X} F,$$

and similarly for Y. An isomorphism $M \xrightarrow{\sim} N$ induces an isomorphism of F-vector spaces $M/\mathfrak{m}M \xrightarrow{\sim} N/\mathfrak{m}N$, which by the above isomorphisms have bases of cardinality $|X|$ and $|Y|$ respectively. Since Y is finite, Theorem 5.5.28 tells us that X must be finite of order $|Y|$. □

The following is immediate.

COROLLARY 5.7.20. *Let R be a commutative ring, and let M be a free R-module on a set of n elements. Then every basis of M has n elements.*

By Theorem 5.7.22, we may make the following definition.

DEFINITION 5.7.21. The *rank* of a free module M over a commutative ring R is the unique $n \ge 0$ such that $M \cong R^n$ if it exists. Otherwise, M is said to have infinite rank.

For an integral domain, we can do somewhat better with a bit of work. In fact, the following result does not require this assumption, but the proof we give does.

THEOREM 5.7.22. *Let R be an integral domain. Let M be a free R-module on a set of n elements, and let Y be a subset of M. Then:*

i. if Y generates M, then Y has at least n elements,

ii. if Y is R-linearly independent, then Y has at most n elements, and

iii. Y is a basis if and only if it generates M and has exactly n elements.

Moreover, a free module on an infinite set cannot be generated by a finite set of elements.

PROOF. Suppose that M is free on n elements. A choice of basis defines an isomorphism $M \xrightarrow{\sim} R^n$ of R-modules, so we may assume that $M = R^n$. Note that R^n is contained in the $Q(R)$-module $Q(R)^n$ via the canonical inclusion, and any generating set Y of R^n spans $Q(R)^n$. But by Theorems 5.5.28 and 5.5.27, this forces Y to have at least n elements. If Y has n elements, then Y would similarly be a basis of $Q(R)^n$. So, if we had $\sum_{i=1}^n c_i y_i = 0$ for some $c_i \in R$ and distinct $y_i \in Y$, then each $c_i = 0$, which means that Y is an R-basis of R^n.

On the other hand, if Y has more than n elements, then by Theorem 5.5.25, the set Y cannot be linearly independent in $Q(R)^n$. That is, there exist $\alpha_i \in Q(R)^n$ and distinct $y_i \in Y$ for $1 \le i \le m$ and $m \ge 1$ with $\sum_{i=1}^m \alpha_i y_i = 0$ and not all $\alpha_i = 0$. For each i, write $\alpha_i = c_i d_i^{-1}$ with $c_i, d_i \in R$ and $d_i \neq 0$. Taking d to be the product of the d_i, we then have $a_i = d\alpha_i \in R$ and not all $a_i = 0$. Since $\sum_{i=1}^m a_i y_i = 0$, it follows that Y is not a basis.

Finally, if N is a free module on an infinite set X, then $N \cong \bigoplus_{x \in X} R$, and so we take N to be the latter module. We then have that $\bigoplus_{x \in X} Q(R)$ is a $Q(R)$-vector space with an infinite basis. But then Theorem 5.5.28 tells us that every basis is infinite, which by Theorem 5.5.27 tells us that a finite set cannot span. □

REMARK 5.7.23. The full analogues of Theorems 5.5.25 and 5.5.27 do not hold for modules over arbitrary rings, over even abelian groups. That is, take the free $\mathbb{Z}$-module $\mathbb{Z}$. The set $\{2\}$ does not span it and is not contained in a basis of $\mathbb{Z}$, and the set $\{2,3\}$ does span it and does not contain a basis.

EXAMPLE 5.7.24. The polynomial ring $R[x]$ is a free R-module on the basis $\{x^i \mid i \in \mathbb{Z}_{\ge 0}\}$.

REMARK 5.7.25. Consider the ideal $I = (2,x)$ of $\mathbb{Z}[x]$. It is not a free $\mathbb{Z}[x]$-module. To see this, first note that it is not a principal ideal so cannot be generated by a single element. As I can be generated by the two elements 2 and x, if I were free, then it would follow from Theorem 5.7.22 that $\{2,x\}$ would be a basis for I. On the other hand, $x \cdot 2 - 2 \cdot x = 0$, which would contradict Proposition 5.7.14.

PROPOSITION 5.7.26. *Let M be an R-module, and let $\pi\colon M \to F$ be a surjective R-module homomorphism, where F is R-free. Then there exists an injective R-module homomorphism $\iota\colon F \to M$ such that $\pi \circ \iota = \mathrm{id}_F$. Moreover, we have $M = \ker(\pi) \oplus \iota(F)$.*

PROOF. Let X be an R-basis of F, and for each $x \in X$, choose $m_x \in M$ with $\pi(m_x) = x$. We take $\iota\colon F \to M$ to be the unique R-module homomorphism with $\iota(x) = m_x$ for all $x \in X$, which exists as F is free. Then $\pi \circ \iota(x) = x$ for all $x \in X$, so $\pi \circ \iota = \mathrm{id}_F$ by uniqueness, and ι must be injective.

Finally, let $A = \ker \pi$. Note that any $m \in M$ satisfies $m - \iota \circ \pi(m) \in A$, so $M = A + \iota(F)$. If $m \in A \cap \iota(F)$, then $m = \iota(n)$ for some $n \in F$ and $n = \pi \circ \iota(n) = \pi(m) = 0$, so $m = 0$. In other words, we have $M = A \oplus \iota(F)$. □

In particular, every free quotient of an R-module M is isomorphic to a direct summand of M.

5.8. Matrix representations

We work in this section with (nonzero) homomorphisms of free modules over a ring R. Most of the time, the case of interest is that of linear transformations of vector spaces over fields, but there is no additional restriction caused by working is full generality.

LEMMA 5.8.1. *Let R be a ring. Let $A \in M_{mn}(R)$ be a matrix for some $m,n \geq 1$. Then there is a unique R-module homomorphism $T\colon R^n \to R^m$ satisfying $T(v) = Av$ for all $v \in R^n$, where Av is matrix multiplication, viewing elements of R^m and R^n as column vectors.*

PROOF. Define $T(e_j) = \sum_{i=1}^m a_{ij} f_i$, where e_j (resp., f_i) is the jth (resp., ith) standard basis element of R^n (resp., R^m). If $v = \sum_{j=1}^n c_j e_j$ for some $c_j \in F$ with $1 \leq j \leq n$, then

$$T(v) = \sum_{j=1}^n c_j T(e_j) = \sum_{i=1}^m \Big(\sum_{j=1}^n a_{ij} c_j \Big) f_i = Av.$$

The uniqueness follows from the fact that R^n is free, so any R-module homomorphism from it is determined by its values on a basis □

DEFINITION 5.8.2. An *ordered basis* is a basis of a free R-module together with a total ordering on the basis.

REMARK 5.8.3. We refer to a finite (ordered) basis on a free R-module as a set $\{v_1, v_2, \ldots, v_n\}$ and take this implicitly to mean that the set has cardinality n and that the basis is ordered in the listed order (i.e., by the ordering $v_i \leq v_{i+1}$ for all $1 \leq i < n$).

EXAMPLE 5.8.4. The standard basis $\{e_1, e_2, \ldots, e_n\}$ on R^n is ordered in the order of positions of the nonzero coordinate of its elements.

NOTATION 5.8.5. If $B = \{v_1, v_2, \ldots, v_n\}$ is an ordered basis of a free R-module V, then we let $\varphi_B\colon R^n \to V$ denote the R-module isomorphism satisfying $\varphi_B(e_i) = v_i$ for all i.

Given ordered bases of free R-modules V and W, an R-module homomorphism $T\colon V \to W$ can be described by a matrix.

DEFINITION 5.8.6. Let V and W be free modules over a ring R with ordered bases $B = \{v_1, v_2, \ldots, v_n\}$ and $C = \{w_1, w_2, \ldots, w_m\}$, respectively. Let $T\colon V \to W$ be an R-module homomorphism. We say that a matrix $A = (a_{ij}) \in M_{nm}(R)$ *represents T with respect to the bases B and C* if

$$T(v_j) = \sum_{i=1}^m a_{ij} w_i$$

for all $1 \leq j \leq n$.

REMARK 5.8.7. Given ordered bases $B = \{v_1, \ldots, v_n\}$ of a free module V and $C = \{w_1, \ldots, w_m\}$ of a free module W, the composition

$$\varphi_C^{-1} \circ T \circ \varphi_B\colon R^n \xrightarrow{\varphi_B} V \xrightarrow{T} W \xrightarrow{\varphi_C^{-1}} R^m,$$

is given by multiplication by a matrix A by Lemma 5.8.1. This A is the matrix representing T with respect to B and C.

TERMINOLOGY 5.8.8. Let V be a free R-module with finite basis B, and let $T\colon V \to V$ be an R-module homomorphism. We say say that a matrix A *represents* T *with respect to* B if A represents T with respect to B and B. If $V = R^n$ and B is the standard basis, we simply say that A *represents* T.

LEMMA 5.8.9. *Let $T'\colon U \to V$ and $T\colon V \to W$ be homomorphisms of finite rank free R-modules. Let B, C, and D be bases of U, V, and W, respectively. Suppose that A' represents T' with respect to B and C and that A represents T with respect to C and D. Then AA' represents $T \circ T'\colon U \to W$ with respect to B and D.*

PROOF. We have that A represents $\varphi_D^{-1} \circ T \circ \varphi_C$ and A' represents $\varphi_C^{-1} \circ T' \circ \varphi_B$. In other words, the maps are left multiplication by the corresponding matrices. The map

$$\varphi_D^{-1} \circ T \circ T' \circ \varphi_B = (\varphi_D^{-1} \circ T \circ \varphi_C) \circ (\varphi_C^{-1} \circ T' \circ \varphi_B),$$

is then left multiplication by AA', which is to say that it is represented by AA'. □

DEFINITION 5.8.10. Let $B = \{v_1, \dots, v_n\}$ and $B' = \{v'_1, \dots, v'_n\}$ be ordered bases of a free R-module V. The *change-of-basis matrix* from B to B' is the matrix $Q_{B,B'} = (q_{ij})$ that represents the R-module homomorphism $T_{B,B'}\colon V \to V$ with $T_{B,B'}(v_i) = v'_i$ for $1 \le i \le n$ with respect to B.

REMARK 5.8.11. If $v'_j = \sum_{i=1}^n q_{ij} v_i$ for all i, then the change-of-basis matrix $Q_{B,B'}$ of Definition 5.8.10 is the matrix (q_{ij}). It is invertible, and $Q_{B',B} = Q_{B,B'}^{-1}$.

REMARK 5.8.12. Let V be free of rank n with bases B and B'. By definition, the change-of-basis matrix $Q_{B,B'}$ represents $\varphi_B^{-1} \circ T_{B,B'} \circ \varphi_B$. On the other hand, we also have that that $\varphi_{B'} = T_{B,B'} \circ \varphi_B$. Thus, see that

$$\varphi_B^{-1} \circ \varphi_{B'} = \varphi_B^{-1} \circ T_{B,B'} \circ \varphi_B,$$

is represented by $Q_{B,B'}$.

THEOREM 5.8.13 (Change of basis theorem). *Let $T\colon V \to W$ be a linear transformation of free R-modules of finite rank. Let B and B' be ordered bases of V and C and C' be ordered bases of W. If A is the matrix representing T with respect to B and C, then $Q_{C,C'}^{-1} A Q_{B,B'}$ is the matrix representing T with respect to B' and C'.*

PROOF. We have that A represents $\varphi_C^{-1} \circ T \circ \varphi_B$, and we wish to compute the matrix representing $\varphi_{C'}^{-1} \circ T \circ \varphi_{B'}$. We have

$$\varphi_{C'}^{-1} \circ T \circ \varphi_{B'} = (\varphi_{C'}^{-1} \circ \varphi_C) \circ (\varphi_C^{-1} \circ T \circ \varphi_B) \circ (\varphi_B^{-1} \circ \varphi_{B'}),$$

and these three matrices are represented by $Q_{C,C'}^{-1}$, A, and $Q_{B,B'}$, respectively. □

CHAPTER 6

Field theory and Galois theory

6.1. Extension fields

DEFINITION 6.1.1. A field E is an *extension field* (or *extension*) of F if F is a subfield of E. We write E/F (which reads "E over F") to denote that E is an extension field of F, and we say that E/F is a *field extension*, or an *extension of fields*.

EXAMPLES 6.1.2. We have that $\mathbb{R}$ is an extension field of $\mathbb{Q}$, and $\mathbb{C}$ is an extension of both $\mathbb{Q}$ and $\mathbb{R}$. We have that $\mathbb{Q}(i)$ is an extension of $\mathbb{Q}$ of which $\mathbb{C}$, but not $\mathbb{R}$, is an extension field.

We will often have cause to deal with the field $\mathbb{Z}/p\mathbb{Z}$, where p is a prime. When we think of $\mathbb{Z}/p\mathbb{Z}$ as a field, we make a change of notation.

DEFINITION 6.1.3. For a prime p, the *field of p elements*, $\mathbb{F}_p$, is $\mathbb{Z}/p\mathbb{Z}$.

LEMMA 6.1.4. *Let F be a field. If F has characteristic* 0*, then F is an extension of* $\mathbb{Q}$*. If F has characteristic equal to a prime p, then F is an extension of* $\mathbb{F}_p$*.*

PROOF. If F has characteristic 0, we define $\iota\colon \mathbb{Q} \to F$ by $\iota(ab^{-1}) = (a \cdot 1) \cdot (b \cdot 1)^{-1}$, where $a, b \in \mathbb{Z}$ and $b \neq 0$. Since ι is a ring homomorphism and $\mathbb{Q}$ is a field, it is injective, so $\mathbb{Q}$ sits isomorphically inside F. If F has characteristic p, then we define $\iota\colon \mathbb{F}_p \to F$ by the same equation, where now $a, b \in \mathbb{Z}$ and $b \not\equiv 0 \bmod p$. Since F has characteristic p, this is a ring homomorphism, and again it is injective. □

DEFINITION 6.1.5. An *intermediate field* of a *field extension* E/F is a subfield E' of E containing F. The extension E'/F is said to be a *subextension* of F in E.

DEFINITION 6.1.6. The *ground field* (or *base field*) of a field extension E/F is the field F.

DEFINITION 6.1.7. Let E/F be a field extension. Let $A \subset E$. The *field generated over F by the set A* (or its elements) is the smallest subfield K of E containing F and A, often denoted $F(A)$. We say that the elements of A *generate* K as an extension of F and that K is given by *adjoining* the elements of A to F.

NOTATION 6.1.8. Let E/F be a field extension and $\alpha_1, \alpha_2, \dots, \alpha_n \in E$ for some $n \geq 0$. We write $F(\alpha_1, \alpha_2, \dots, \alpha_n)$ for the subfield of E generated by the set $\alpha_1, \alpha_2, \dots, \alpha_n$ over F.

REMARK 6.1.9. One often says "F adjoin α" to refer to a field $F(\alpha)$.

REMARK 6.1.10. Note that the field generated over F by a set of elements A of E is well-defined, equal to the intersection of all subfields of E containing both F and A.

REMARK 6.1.11. Note that we distinguish between the field $F(x_1,x_2,\dots,x_n)$ of rational functions, where $x_1,x_2,\dots,x_n$ are indeterminates, and $F(\alpha_1,\alpha_2,\dots,\alpha_n)$, where $\alpha_1,\alpha_2,\dots,\alpha_n$ are elements of an extension field of F. These fields can be quite different. However, in that $F(\alpha_1,\alpha_2,\dots,\alpha_n)$ is the quotient field of $F[\alpha_1,\alpha_2,\dots,\alpha_n]$, every element of $F(\alpha_1,\alpha_2,\dots,\alpha_n)$ is a rational function in the elements α_i with $1 \le i \le n$.

EXAMPLE 6.1.12. The fields $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(i)$ are extension fields of $\mathbb{Q}$ inside $\mathbb{R}$ and $\mathbb{C}$, respectively.

PROPOSITION 6.1.13. *Let E/F be a field extension, and let $\alpha \in E$. Then $F(\alpha)$ is isomorphic to the quotient field of $F[\alpha]$.*

PROOF. Since $F[\alpha]$ is the smallest subring of E containing F and α and $F(\alpha)$ is the smallest subfield of E containing F and α, inclusion provides an injective homomorphism

$$\iota\colon F[\alpha] \to F(\alpha).$$

Since $F(\alpha)$ is a field, ι induces an injective map $Q(\iota)\colon Q(F[\alpha]) \to F(\alpha)$. Since the image of $Q(\iota)$ is a subfield of $F(\alpha)$ containing F and α and $F(\alpha)$ is the smallest such field in E, we have that $Q(\alpha)$ is surjective as well. □

In many cases, an extension field generated by an element is actually equal to the ring generated by the element. We see this holds in a couple of simple examples.

EXAMPLE 6.1.14. The fields $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(i)$ equal $\mathbb{Q}[\sqrt{2}]$ and $\mathbb{Q}[i]$ as subrings of $\mathbb{R}$ and $\mathbb{C}$ respectively. E.g., the elements of $\mathbb{Q}(\sqrt{2})$ all may be written in the form $a+b\sqrt{2}$ for some $a,b \in \mathbb{Q}$.

The key in this example is that $\sqrt{2}$ and i are roots of polynomials with coefficients in $\mathbb{Q}$. Let us examine this further.

THEOREM 6.1.15. *Let E be an extension field of a field F, and let $f \in F[x]$ be an irreducible polynomial that has a root $\alpha \in E$. Then the evaluation map $e_\alpha\colon F[x] \to F(\alpha)$ given by $e_\alpha(f) = f(\alpha)$ induces an isomorphism*

$$\overline{e_\alpha}\colon F[x]/(f) \xrightarrow{\sim} F(\alpha)$$

of fields such that $\overline{e_\alpha}(a) = a$ for all $a \in F$.

PROOF. First, note that we have the inclusion map $F \to F[x]$ and the quotient map $F[x]/(f)$, inducing a nonzero, and hence injective, map of fields $F \to F[x]/(f)$. This allows us to view F as a subfield of $F[x]/(f)$. The map e_α has kernel containing f, and if $e_\alpha(g) = g(\alpha) = 0$ for some $g \in F[x]$, then g also has α as a root. Since any GCD of f and g will then have α as a root, we have that the GCD of f and g is (f), and in particular, f divides g, so $g \in (f)$. Therefore, $\overline{e_\alpha}$ is injective. Since the image of $\overline{e_\alpha}$ is a field containing F and α, it must then be equal to $F(\alpha)$. □

EXAMPLE 6.1.16. The field $\mathbb{Q}(i)$ is isomorphic to the quotient ring $\mathbb{Q}[x]/(x^2+1)$.

We now obtain the following theorem as a corollary.

THEOREM 6.1.17 (Kronecker). *Let F be a field, $f \in F[x]$ a polynomial. Then there exists a field extension E of F and an element $\alpha \in E$ such that $f(\alpha) = 0$.*

PROOF. First, we may assume that f is irreducible by replacing f by an irreducible polynomial dividing it which has α as a root. We then set $E = F[x]/(f)$, which is a field. Let $\alpha = x + (f)$. Then $f(\alpha)$ is the image of f in E, and so $f(\alpha) = 0$. □

DEFINITION 6.1.18. Let E/F be a field extension. A nonconstant polynomial $f \in F[x]$ is said to *split* (or *factor completely*) in E if it can be written as a product of linear polynomials in $E[x]$.

DEFINITION 6.1.19. Let F be a field. A *splitting field* E for $f \in F[x]$ over F is an extension of F such that f splits in E but not any proper subextension of F in E.

EXAMPLES 6.1.20.

a. The field $\mathbb{Q}(\sqrt{2})$ is the splitting field of $x^2 - 2$, since it contains both of its roots. It is also the splitting field of $(x-a)^2 - 2b^2$ for any $a, b \in \mathbb{Q}$.

b. The field $\mathbb{Q}(\sqrt[3]{2})$ is not the splitting field of $x^3 - 2$, since it contains $\sqrt[3]{2}$ but not its other two roots. On the other hand, if $\omega \in \mathbb{C}$ is a primitive cube root of unity, then $\mathbb{Q}(\sqrt[3]{2}, \omega\sqrt[3]{2}, \omega^2\sqrt[3]{2})$ is a splitting field for $x^3 - 2$ inside $\mathbb{C}$. This field may be written more simply as $\mathbb{Q}(\omega, \sqrt[3]{2})$.

As a corollary of Kronecker's theorem, we have the following.

COROLLARY 6.1.21. *Let F be a field, and let $f \in F[x]$. Then there exists a splitting field for f over F.*

PROOF. Let $n = \deg f$. The result is clearly true for $n = 1$. Set $K = F[x]/(f)$. Then f has a root α in K by Kronecker's theorem. Set $g(x) = (x-\alpha)^{-1} f(x) \in E[x]$. Then there exists a splitting field E of g over K which is generated by the roots of g over F by induction. This E is a splitting field of f over F, since it is generated by the roots of f. □

We next distinguish two types of elements of extension fields of F: those that are roots of polynomials and those that are not.

DEFINITION 6.1.22. Let E/F be a field extension. An element $\alpha \in E$ is called *algebraic* over F if there exists a nonzero $f \in F[x]$ such that $f(\alpha) = 0$. Otherwise, α is said to be *transcendental* over F.

When speaking of elements of extensions of $\mathbb{Q}$, we speak simply of algebraic and transcendental numbers.

DEFINITION 6.1.23. An element of $\mathbb{C}$ is said to be an *algebraic number* if it is algebraic over $\mathbb{Q}$ and a *transcendental number* if it is transcendental over $\mathbb{Q}$.

EXAMPLES 6.1.24. The number $\sqrt{2}$ is an algebraic number, since it is a root of $x^2 - 2$. Similarly, i is algebraic, being a root of $x^2 + 1$. However, the real number π is transcendental, and the real number e such that $\log e = 1$ is transcendental as well. We do not prove the latter two facts here.

EXAMPLE 6.1.25. A real number given by repeated square roots

$$\sqrt{a_1 + \sqrt{a_2 + \cdots + \sqrt{a_n}}}$$

with the a_i positive rational numbers is algebraic: it is a root of

$$(\cdots((x^2-a_1)^2-a_2)^2\cdots)^2-a_n.$$

EXAMPLE 6.1.26. If F is a field and $\alpha \in F$, then α is algebraic over F, being a root of $x-\alpha$.

Note the following.

PROPOSITION 6.1.27. *Let E/F be a field extension, and let $\alpha \in E$. Then α is transcendental over F if and only if the evaluation homomorphism $e_\alpha \colon F[x] \to E$ is injective.*

PROOF. By definition, $\alpha \in E$ is transcendental if and only if $g(\alpha) \neq 0$ for every $g \in F[x]$ that is nonzero. But $g(\alpha) = e_\alpha(g)$, so we are done. □

This allows us to give the prototypical example of a transcendental element.

COROLLARY 6.1.28. *Let F be a field. The element x of the field $F(x)$ of rational functions of F is transcendental over F.*

PROOF. Let y be an indeterminate. Consider $e_\alpha \colon F[y] \to F(x)$ given by $e_\alpha(g) = g(x)$. We have that the polynomial $g(x)$ is zero in $F(x)$ if and only if it is zero in $F[x]$, and therefore if and only if $g = 0$ in $F[y]$. □

THEOREM 6.1.29. *Let E/F be a field extension, and let $\alpha \in E$ be algebraic over F. Then there exists a unique monic, irreducible polynomial $f \in F[x]$ such that $f(\alpha) = 0$.*

PROOF. Since α is algebraic over F, there exists a polynomial $g \in F[x]$ such that $g(\alpha) = 0$. Since g factors as a product of irreducible polynomials, and E is in particular an integral domain, one of the irreducible factors must have α as a root, and by multiplying it by a constant, we may take it to be monic. So suppose that f is a monic irreducible polynomial in $F[x]$ with $f(\alpha) = 0$. Without loss of generality, we may assume that $\deg f$ is minimal among all such polynomials. If $f' \in F[x]$ satisfies $f'(\alpha) = 0$, then the division algorithm provides $q, r \in F[x]$ with $r = f' - qf$ and either $\deg r < \deg f$ or $r = 0$. Since $r(\alpha) = 0$, we must have $r = 0$, but then $f' = qf$, so f divides f'. If f' were monic and irreducible, this would force $f' = f$, as desired. □

DEFINITION 6.1.30. Let E/F be a field extension, and let $\alpha \in E$ be algebraic over F. The *minimal polynomial* of α over F is the unique monic irreducible polynomial in $F[x]$ which has α as a root.

EXAMPLES 6.1.31.

a. If F is a field and $\alpha \in F$, then $x - \alpha$ is the minimal polynomial of α over F.

b. The polynomial $x^2 + 1$ is the minimal polynomial of i over $\mathbb{Q}$.

6.2. Finite extensions

REMARK 6.2.1. If E/F is an extension of fields, then E is an F-vector space via the restriction of the multiplication in E to a map $F \times E \to E$, which is then given by $a \cdot \alpha = a\alpha$ for $a \in F$ and $\alpha \in E$. Moreover, E actually contains F, so F is a F-subspace of E.

DEFINITION 6.2.2. An extension E/F of fields is *finite* if E is a finite-dimensional field extension of F. Otherwise, E/F is said to be an *infinite* extension.

EXAMPLE 6.2.3. The field $\mathbb{Q}(\sqrt{2})$ is an extension of $\mathbb{Q}$ with basis $\{1, \sqrt{2}\}$ and hence is a basis of $\mathbb{Q}(\sqrt{2})$ over $\mathbb{Q}$.

EXAMPLE 6.2.4. The field $F(x)$ of rational functions over a field F is infinite. More generally, if E/F is an extension that contains an element that is transcendental over F, then E is an infinite extension of F.

DEFINITION 6.2.5. The *degree* $[E : F]$ of a finite extension E of a field F is defined to be the dimension $\dim_F E$ of E as a vector space over F. If E/F is an infinite extension, we say that the degree of E over F is *infinite*.

EXAMPLE 6.2.6. The degree $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}]$ is 2, as $\sqrt{2}$ has minimal polynomial $x^2 - 2$. The set $\{1, \sqrt{2}\}$ forms a basis of $\mathbb{Q}(\sqrt{2})$ as a $\mathbb{Q}$-vector space.

The following is essential to our studies.

THEOREM 6.2.7. *Let F be a field, and let $f \in F[x]$ be an irreducible polynomial of degree n. Then the field $F[x]/(f)$ has degree n over F.*

PROOF. Since (f) contains only multiples of f, it contains no nontrivial linear combinations of the monomials x^i with $0 \le i \le n-1$. In other words, the $x^i + (f)$ with $0 \le i \le n-1$ are linearly independent over F. On the other hand, if $g \in F[x]$, then $g = qf + r$ with $q, f \in F[x]$ and $\deg r < n$, so $g + (f) = r + (f)$, and therefore $g + (f)$ may be written as the image in the quotient of a linear combination of the monomials x^i with $0 \le i \le n-1$. That is, the elements $x^i + (f)$ with $0 \le i \le n-1$ form a basis of $F[x]/(f)$. □

The proof of Theorem 6.2.7, when taken together with Theorem 6.1.15, yields the following.

COROLLARY 6.2.8. *Let E/F be a field extension, and let $\alpha \in E$ be algebraic over F. Let n be the degree of the minimal polynomial of F. Then $[F(\alpha) : F] = n$, and $\{1, \alpha, \dots, \alpha^{n-1}\}$ is a basis of $F(\alpha)$ over F.*

PROPOSITION 6.2.9. *If E/F is a finite extension and $\alpha \in E$, then α is algebraic over F.*

PROOF. Since $[E : F]$ is finite, there is an $n \ge 1$ such that the set $\{1, \alpha, \dots, \alpha^n\}$ is F-linearly dependent. We then have

$$\sum_{i=0}^{n} c_i \alpha^i = 0$$

for some $c_i \in F$ with $0 \le i \le n$. Setting $f = \sum_{i=0}^{n} c_i x^i \in F[x]$, we see that $f(\alpha) = 0$, so α is algebraic. □

COROLLARY 6.2.10. *Every finite extension E of a field F has the form $E = F(\alpha_1, \alpha_2, \dots, \alpha_n)$ for some algebraic $\alpha_i \in E$ for $1 \le i \le n$.*

PROOF. One may simply take $\{\alpha_i \mid 1 \le i \le n\}$ to be a basis of E over F. Since each $\alpha_i \in E$, we have $F(\alpha_1, \alpha_2, \dots, \alpha_n) \subseteq E$, and since every element in E is a linear combination of the α_i, we have the opposite containment. □

The following theorem, while stated for arbitrary field extensions, has a number of applications to finite extensions.

THEOREM 6.2.11. *Let E be an extension of a field F, and let K be an extension of E. If A is a basis of E over F and B is a basis of K over E, then*

$$AB = \{\alpha\beta \mid \alpha \in A, \beta \in B\}$$

is a basis of K over F, and the map $A \times B \to AB$ given by multiplication in K is a bijection.

PROOF. We first show that AB spans K. By definition of B, any $\gamma \in K$ can be written as

$$\gamma = \sum_{j=1}^{m} c_j \beta_j$$

with $c_j \in E$ and $\beta_j \in B$ for $1 \le j \le m$ and some $m \ge 1$. Each c_j is in the F-span of some finite subset of A. By taking the union of these subsets, we see that there is a single finite subset of A such that every c_j with $1 \le j \le m$ is in its span. That is, we may write

$$c_j = \sum_{i=1}^{n} d_{ij} \alpha_i$$

for some $d_{ij} \in F$ and $\alpha_i \in E$ for $1 \le i \le n$ and some $n \ge 1$. Plugging in, we obtain

$$\gamma = \sum_{i=1}^{n} \sum_{j=1}^{m} d_{ij} \alpha_i \beta_j,$$

so the set AB spans K over F.

Now, if some F-linear combination of the elements of AB equals zero, then in particular (by throwing in terms with zero coefficients if needed) we may write

$$\sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} \alpha_i \beta_j = 0$$

for some $\alpha_i \in A$, $\beta_j \in B$, and $a_{ij} \in F$ for $1 \le i \le n$ and $1 \le j \le m$, for some m and n. Since the β_j are E-linearly independent, this implies that

$$\sum_{i=1}^{n} a_{ij} \alpha_i = 0$$

for all $1 \le j \le m$. Since the α_i are F-linearly independent, we then have that $a_{ij} = 0$ for all i and j. Therefore, the set AB is a basis of K over F.

Note that we may also conclude that the surjection $A \times B \to AB$ given by multiplication in K is injective. If it were not, then we would have two distinct pairs $(\alpha, \beta), (\alpha', \beta') \in A \times B$ such that $\alpha\beta - \alpha'\beta' = 0$, contrary to what we have shown. □

Theorem 6.2.11 has the following almost immediate corollary.

COROLLARY 6.2.12. *Let E be a finite extension of a field F, and let K be a finite extension of E. Then K/F is a finite extension, and we have*

$$[K : F] = [K : E][E : F].$$

PROOF. Theorem 6.2.11 tells us that any basis of K over F has $[K:E][E:F]$ elements, hence the result. □

This corollary has in turn the following two corollaries.

COROLLARY 6.2.13. *Let E be a finite extension of a field F, and let K be a finite extension of E. Then $[K:E]$ and $[E:F]$ divide $[K:F]$.*

COROLLARY 6.2.14. *Let K/F be a finite extension, and let E be a subfield of K containing F. Then K/E and E/F are finite extensions.*

EXAMPLE 6.2.15. By Corollary 6.2.12, we have

$$[\mathbb{Q}(i,\sqrt{2}):\mathbb{Q}] = [\mathbb{Q}(i,\sqrt{2}):\mathbb{Q}(\sqrt{2})][\mathbb{Q}(\sqrt{2}):\mathbb{Q}],$$

and since $i \notin \mathbb{Q}(\sqrt{2})$, we have that x^2+1 is irreducible in $\mathbb{Q}(\sqrt{2})[x]$, so $[\mathbb{Q}(i,\sqrt{2}):\mathbb{Q}(\sqrt{2})] = 2$. Therefore, $[\mathbb{Q}(i,\sqrt{2}):\mathbb{Q}] = 4$.

We give another corollary of Corollary 6.2.12 that is a converse to Corollary 6.2.10.

COROLLARY 6.2.16. *Let K/F be a field extension, and let $\alpha_1,\alpha_2,\ldots,\alpha_n \in K$ be algebraic. Then $F(\alpha_1,\alpha_2,\ldots,\alpha_n)$ is a finite extension of F.*

PROOF. The corollary is true for $n=1$ by definition of an algebraic element. Suppose by induction we know it for $n-1$, and let $E = F(\alpha_1,\alpha_2,\ldots,\alpha_{n-1})$, which is a finite extension of F by induction. Note that α_n is algebraic over E in that it is algebraic over F. Since $K = E(\alpha_n)$, we therefore have that K is a finite extension of E. That K/F is a finite extension now follows from Corollary 6.2.12. □

DEFINITION 6.2.17. A field extension E/F is said to be *algebraic* if every element of E is algebraic over F. Otherwise, E/F is said to be a *transcendental* extension.

PROPOSITION 6.2.18. *Every finite extension is algebraic.*

PROOF. If $\alpha \in E$, then $F(\alpha) \subseteq E$, so $F(\alpha)/F$ is finite. Hence, α is algebraic over F. □

In fact, we can do better.

PROPOSITION 6.2.19. *Let E be an intermediate field in a field extension K/F. Then K/F is algebraic if and only if both K/E and E/F are algebraic.*

PROOF. Suppose that K/E and E/F are algebraic. Let $\alpha \in K$, and let $f = \sum_{i=0}^{n} a_i x^i \in E[x]$ be its minimal polynomial over E. Since E/F is algebraic, the field $E_f = F(a_1,\ldots,a_n)$ is finite over of F, and therefore so is $E_f(\alpha)$. In particular, α is algebraic over F, and therefore K/F is algebraic. The other direction is immediate. □

REMARK 6.2.20. A transcendental field extension can never be finite.

EXAMPLES 6.2.21.

a. The field $\mathbb{R}$ is a transcendental extension of $\mathbb{Q}$.

b. The field $K = \mathbb{Q}(\sqrt{2},\sqrt[3]{2},\sqrt[4]{2},\sqrt[5]{2},\ldots)$ is an algebraic extension of $\mathbb{Q}$, as the field generated by any finite list of these roots is equal to $\mathbb{Q}(\sqrt[n]{2})$ for $n \geq 2$, every element of K is contained such a field, and each of these fields is algebraic over $\mathbb{Q}$.

6.3. Composite fields

DEFINITION 6.3.1. Let E_1 and E_2 be subfields of a field K. The *compositum*, or *composite field*, E_1E_2 of E_1 and E_2 is the smallest subfield of K containing both E_1 and E_2.

REMARK 6.3.2. The compositum E_1E_2 of subfields E_1 and E_2 of a field K is the intersection of all subfields of K containing both E_1 and E_2.

EXAMPLE 6.3.3. Let K/F be a field extension, and let $\alpha, \beta \in K$. Then

$$F(\alpha,\beta) = F(\alpha)F(\beta).$$

More generally, if E is any subfield of K containing F, then

$$EF(\alpha) = E(\alpha).$$

We prove the following in the case of finite extensions. Note, though, that this finiteness is not needed, as seen through Corollary 6.3.11 below.

PROPOSITION 6.3.4. *Let E_1 and E_2 be finite extensions of a field F contained in a field K. Suppose that A and B are bases of E_1 and E_2 as F-vector spaces, respectively. Then E_1E_2 is spanned by the set AB.*

PROOF. Set $m = [E_1 : F]$ and $n = [E_2 : F]$, and let $A = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$ and $B = \{\beta_1, \beta_2, \dots, \beta_n\}$. Clearly, we have

$$E_1E_2 = F(\alpha_1, \alpha_2, \dots, \alpha_m, \beta_1, \beta_2, \dots, \beta_n).$$

As the elements of A and B are algebraic, we have $E(\alpha_i) = E[\alpha_i]$ and $E(\beta_i) = E[\beta_i]$ for any field E containing F, for all i and j. We then see by a simple recursion that every element of E_1E_2 may actually be expressed as a polynomial in the elements of A and B with coefficients in F, not just a rational function. However, any monomial in the elements of A lies in E_1, hence may be written as a linear combination of the elements of A. Similarly, any monomial in the elements of B lies in E_2, hence may be written as a linear combination of the elements of B. Therefore, every monomial is the elements of A and B may be written as a product of a linear combination of elements of A with a linear combination of elements of B, which is the a linear combination of elements of AB. Since every polynomial is a linear combination of monomials, we are done. □

COROLLARY 6.3.5. *Let E_1 and E_2 be finite extensions of a field F that are contained in a field K. Then we have*

$$[E_1E_2 : F] \le [E_1 : F][E_2 : F].$$

PROOF. Let A (resp., B) be a basis of E_1 (resp., E_2) over F. Then AB has at most $[E_1 : F][E_2 : F]$ elements and spans E_1E_2 over F. □

COROLLARY 6.3.6. *Let E_1 and E_2 be finite extensions of a field F that are contained in a field K, and suppose that $[E_1 : F]$ and $[E_2 : F]$ are relatively prime. Then we have*

$$[E_1E_2 : F] = [E_1 : F][E_2 : F].$$

PROOF. Both $[E_1 : F]$ and $[E_2 : F]$ divide $[E_1E_2 : F]$, so by their relative primality, their product $[E_1 : F][E_2 : F]$ does as well. So we have $[E_1E_2 : F] \ge [E_1 : F][E_2 : F]$, while Corollary 6.3.5 provides the opposite inequality. □

DEFINITION 6.3.7. Let $n \geq 1$. An nth root of unity is an element of order dividing n in the multiplicative group of a field.

That is, if F is a field, $\zeta \in F$ is an nth root of unity if and only if $\zeta^n = 1$.

EXAMPLE 6.3.8. Let ω be a third root of unity in $\mathbb{C}$ that is not equal to 1. Note that $[\mathbb{Q}(\omega) : \mathbb{Q}] = 2$, since $\omega^2 + \omega + 1 = 0$. Then $\sqrt[3]{2}$ and $\omega\sqrt[3]{2}$ are both cube roots of 2, and we have

$$\mathbb{Q}(\sqrt[3]{2}, \omega\sqrt[3]{2}) = \mathbb{Q}(\omega, \sqrt[3]{2})$$

We then see that

$$[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}][\mathbb{Q}(\omega\sqrt[3]{2}) : \mathbb{Q}] = 9,$$

while

$$[\mathbb{Q}(\omega, \sqrt[3]{2}) : \mathbb{Q}] = [\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}][\mathbb{Q}(\omega) : \mathbb{Q}] = 6$$

by Corollary 6.3.6.

More generally, we may define the compositum of a collection of fields.

DEFINITION 6.3.9. Let $\{E_i \mid i \in I\}$ be a collection of subfields of a field K for some indexing set I. Then the *compositum* of the fields E_i for $i \in I$ is smallest subfield of K containing all E_i.

Let us give an alternate description of the compositum.

LEMMA 6.3.10. *Let $\{E_i \mid i \in I\}$ be a collection of intermediate fields in an extension K/F for some indexing set I. Then the compositum E of the E_i is equal to the union of its subfields $F(\alpha_1, \alpha_2, \ldots, \alpha_n)$, where $n \geq 0$ and each α_j with $1 \leq j \leq n$ is an element of E_i for some $i \in I$.*

PROOF. Clearly the above-described union U is contained in E and contains each E_i. However, we must show that U is a field, hence equal to E. If $a, b \in U$ are nonzero, then $a \in F(\alpha_1, \alpha_2, \ldots, \alpha_n)$ and $b \in F(\beta_1, \beta_2, \ldots, \beta_m)$, where $n, m \geq 0$ and the α_j and β_k are elements of the E_i. Then

$$a - b, ab^{-1} \in F(\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \beta_2, \ldots, \beta_m),$$

and the latter field is a subset of U, so $U = E$. □

We have the following corollary.

COROLLARY 6.3.11. *Let $\{E_i \mid i \in I\}$ be algebraic extensions of a field F that are contained in a field K, where I is an indexing set. Then the compositum E of the fields E_i is an algebraic extension of F.*

PROOF. By Lemma 6.3.10, any $\alpha \in E$ is an element of a subfield $F(\alpha_1, \alpha_2, \ldots, \alpha_n)$ of E, where each $\alpha_j \in E_i$ for some $i \in I$. Since E_i is algebraic, $F(\alpha_j)/F$ is finite for all $1 \leq j \leq n$, and therefore $F(\alpha_1, \alpha_2, \ldots, \alpha_n)/F$ is finite by Corollary 6.3.5. □

6.4. Constructible numbers

In this section, we discuss a classical problem of the ancient Greeks, which we present as a game. The game begins with a line segment of length 1 that has already been drawn on the plane. One is given two tools: a straightedge and a compass. At any step of the game, one can either use the straightedge to draw a line segment or the compass to draw a circle, in ways we will shortly make more specific. The goal of the game is to draw a line segment of a given desired length in a finite number of steps.

At any step, we consider a point to have been marked if it is either the endpoint of an already drawn line segment or the intersection of a drawn line segment or circle with another drawn line segment or circle. The straightedge allows us to draw a line segment between any two marked points and also to extend any previously drawn line segment until it meets any point that has already been drawn on the plane. The compass allows us to draw a circle that contains a given marked point and has as its center any other marked point.

Given these rules, we may now make the following definition.

DEFINITION 6.4.1. A real number α is said to be *constructible* if one can draw a line segment of length $|\alpha|$ in the plane, starting from a line segment of length 1, using a straightedge and compass, in a finite number of steps.

We will denote a line segment between two distinct points A and B in $\mathbb{R}^2$ by $\overline{AB}$. Its length will be denoted by $|\overline{AB}|$. We prove a few preliminary results.

LEMMA 6.4.2. *Suppose that a line segment $\overline{AB}$ has been drawn in the plane.*

a. We may draw a line segment bisecting $\overline{AB}$.

b. We may draw a line segment $\overline{AC}$ perpendicular to $\overline{AB}$.

c. Given a point D in the plane, we may draw a line segment $\overline{DE}$ parallel to $\overline{AB}$.

PROOF. For part a, draw circles with center A and center B, both of radius $|\overline{AB}|$. These intersect at two points, and the line segment between them is perpendicular to $\overline{AB}$ and passes through a midpoint F of that segment.

For part b, by drawing the circle with center A and radius $|\overline{AF}|$, we may mark a point G on the line that contains $|\overline{AB}|$ that is on the opposite side of A from F and is such that $|\overline{AF}| = |\overline{AG}|$. As we have already shown, we may then draw a line segment $\overline{CH}$ bisecting $\overline{FG}$ and passing through A, which provides us with $\overline{AC}$.

For part c, if $\overline{BD}$ is perpendicular to $\overline{AB}$, we set $H = D$. Otherwise, we draw a circle with center D and passing through B. It intersects $\overline{AB}$ in a second point H. We draw a line segment through D bisecting $\overline{BH}$ using part a. We then use part b to draw a perpendicular $\overline{DE}$ to $\overline{BH}$, and it is by definition parallel to $\overline{AB}$. □

We also have the following.

LEMMA 6.4.3. *Suppose we have drawn either a line segment of length α or a circle of radius α in the plane.*

a. We may draw a line segment of length α with any marked point as an endpoint, along any line that contains at least one other marked point.

b. We may draw a circle of radius α with center any marked point.

PROOF. First we note that the two assumptions are equivalent. Given a line segment of length α, we may use its endpoints to draw a circle of radius α. Given a circle of radius α and center A, since we have at least one marked point other than its center in the plane, we can by drawing the line segment from the center to that point mark a point B on the circle. The resulting line segment B then has radius α.

Suppose then that we are given a line segment $\overline{AB}$ of length α and another line segment $\overline{CD}$. Me make two constructions using Lemma 6.4.2. We draw a line segment $\overline{CE}$ parallel to $\overline{AB}$. We draw the line segment $\overline{AC}$ and then the parallel to $\overline{AC}$ passing through B. It intersects the line through C and E at a point F such that $|\overline{AF}| = \alpha$. The circle with center A passing through F then determines a point G on the line through A and C such that $|\overline{AG}| = \alpha$. We thus have both parts. □

We prove the following.

THEOREM 6.4.4. *The set of constructible numbers is a subfield of* $\mathbb{R}$.

PROOF. Suppose that α and β are constructible and positive. Then we may draw a line segment $\overline{AB}$ of length α in the plane, and we may then draw a line segment $\overline{BC}$ of length β along the line defined by $\overline{AB}$. If we do this so that it overlaps with $\overline{AB}$, then we have constructed a line segment $\overline{AC}$ of length $|\alpha - \beta|$.

On the other hand, given $\overline{AB}$ of length α, draw a line segment $\overline{AC}$ of length β that is perpendicular to $\overline{AB}$, and let E be the point on the ray defined by $\overline{AC}$ such that $\overline{AE}$ has length 1. Draw the line segment $\overline{CB}$, and use it to draw a parallel line segment from E to a point D on the ray defined by the segment $\overline{AB}$. We then have that the triangle ABC is similar to the triangle ADE, so

$$|\overline{AD}| = \frac{|\overline{AD}|}{|\overline{AE}|} = \frac{|\overline{AB}|}{|\overline{AC}|} = \frac{\alpha}{\beta}.$$

Therefore, $\alpha\beta^{-1}$ is constructible. □

THEOREM 6.4.5. *The field of constructible numbers consists exactly of the real numbers that can be obtained from* 1 *by applying a finite sequence of the operations of addition, subtraction, multiplication, division (with nonzero denominators), and the taking of square roots (of positive numbers), using numbers already obtained from* 1 *at an earlier point in the sequence.*

PROOF. We first show that the square root of a constructible positive number α is constructible. For this, draw a line segment $\overline{AD}$ of length $1+\alpha$ and mark a point B at distance 1 from A and α from D along the segment. Find the midpoint O of $\overline{AD}$, and draw a circle with center O and radius $|\overline{AO}| = (1+\alpha)/2$. Draw a perpendicular to $\overline{AD}$ at the point B, and let C be a point where it intersects the drawn circle. Then the triangle ABC is similar to the triangle CBD, and therefore we have

$$|\overline{BC}| = \frac{|\overline{BC}|}{|\overline{AB}|} = \frac{|\overline{BD}|}{|\overline{BC}|} = \frac{\alpha}{|\overline{BC}|},$$

and hence $|\overline{BC}| = \sqrt{\alpha}$.

For the converse, we merely give a sketch. Let E be the set (or actually, field) of numbers that can be constructed from 1 using field operations and square roots. Suppose that our initial line segment was between $(0,0)$ and $(1,0)$ on the plane. Suppose that all previously marked points have coordinates in E. These points have been marked as the intersection points of lines and circles, where the lines are determined by previously marked points with E-coordinates and the circles have centers previously marked points with E-coordinates and are chosen to pass through marked points with E-coordinates. Every drawn line thus has the form $ax+by+c=0$ with $a,b,c \in E$, and every drawn circle has the form $x^2+y^2+dx+ey+f=0$ with $d,e,f \in E$. The intersection of two such lines has coordinates obtained by field operations on the coefficients of the two lines in question. The coordinates of the intersection points of a line and a circle coming from the solution of a quadratic equation with coefficients are obtained by field operations on the coefficients of the line and the circle. Finally, the intersection of two circles can be reduced to the latter case by considering a common chord (or tangent line). It follows that any new line segment or circle created with these operations has two marked points in E, and therefore every constructible length lies in E as well. □

Since the square root of a field element defines an extension of degree dividing 2 of the field in which it lies, we have the following.

COROLLARY 6.4.6. *Let α be a constructible number. Then α is an algebraic number, and $[\mathbb{Q}(\alpha):\mathbb{Q}]$ is a power of* 2.

COROLLARY 6.4.7. *The field of constructible numbers is an algebraic extension of* $\mathbb{Q}$.

The ancient Greeks were in particular very concerned with three problems that they could not solve with a straightedge and compass. This was for good reason: they involved constructing line segments of unconstructible length. Yet, the Greeks never managed to prove this, and it was not until the 19th century that proofs were finally given. We list these three problems now.

EXAMPLES 6.4.8.

a. It is impossible to "double the cube." That is, given a line segment, one cannot construct from it a new line segment such that a cube with the new line segment as one of its sides would have twice the volume of a cube with the original line segment as its side. Assuming the initial line segment had a constructible length α, the new line segment would have to have length $\sqrt[3]{2}\alpha$, but then $\sqrt[3]{2}$ would be constructible, yet it defines an extension of degree 3 over $\mathbb{Q}$, in contradiction to Corollary 6.4.6.

b. It is impossible to "square the circle." That is, given a drawn circle, it is impossible to construct a square with the same area. If the circle had radius r, then the square would have side length $\sqrt{\pi}r$, which would mean that $\sqrt{\pi}$ would be constructible, and hence π would be as well, in contradiction to Corollary 6.4.6, since π is transcendental.

c. It is impossible to "trisect all angles." That is, given an arbitrary angle between two drawn line segments with a common endpoint in a plane, it is not always possible to draw a line segment with the same endpoint having an angle with one of the line segments that is a third of the original angle. Note that an initial such angle θ exists if and only if $\cos\theta$ is constructible, as seen by drawing a perpendicular from one line segment at point a distance one from the point of intersection until it intersects the line defined by the other. Therefore, the problem is, given a constructible number $\alpha = \cos\theta$, to show that $\cos(\theta/3)$ is constructible. However, we have a trigonometric identity

$$\cos\theta = 4\cos^3(\theta/3) - 3\cos(\theta/3).$$

Suppose that $\theta = \pi/3$. Then $\cos(\pi/3) = \frac{1}{2}$, and $\cos(\pi/9)$ would be a root of the polynomial $8x^3 - 6x - 1$, which is irreducible over $\mathbb{Q}$ since it is irreducible in $\mathbb{Z}[x]$. (It has no roots, even modulo 2.) But then $\cos(\pi/9)$ would define a degree 3 extension of $\mathbb{Q}$, contradicting Corollary 6.4.6 again.

6.5. Finite fields

In this section, we classify all finite fields, which is to say, fields of finite order.

NOTATION 6.5.1. We use $\mathbb{F}_p$ to denote $\mathbb{Z}/p\mathbb{Z}$ when we consider it as a field.

PROPOSITION 6.5.2. *Every finite field contains p^n elements for some $n \geq 1$.*

PROOF. Let F be a finite field. Since it is finite, it has characteristic p for some prime number p, which means that it contains the field $\mathbb{F}_p$, and moreover is a finite dimensional vector space over $\mathbb{F}_p$. Therefore, F has a finite $\mathbb{Z}/p\mathbb{Z}$-basis $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$, so that the elements of F are exactly the elements $c_1\alpha_1 + c_2\alpha_2 + \cdots + c_n\alpha_n$ with $c_1, c_2, \dots, c_n \in \mathbb{Z}/p\mathbb{Z}$. We therefore have $|F| = p^n$. □

DEFINITION 6.5.3. Let F be a field and n be a positive integer. The group $\mu_n(F)$ of nth roots of unity in F is the subgroup of $F^\times$ with elements the nth roots of 1 in F.

LEMMA 6.5.4. *Let F be a field and n be a positive integer. Then $\mu_n(F)$ is a cyclic group of order dividing n.*

PROOF. Every element in $\mu_n(F)$ has order dividing n. Let m be the exponent of $\mu_n(F)$. Then every element of $\mu_n(F)$ is an mth root of unity, so is a root of $x^m - 1$, and hence the order of $\mu_n(F)$ is at most m. On the other hand, since m is the exponent, the classification of finite abelian groups tells us that $\mu_n(F)$ contains an element of order m, so therefore $\mu_n(F)$ is cyclic of order m, which divides n. □

PROPOSITION 6.5.5. *Let F be a finite field of order p^n for some prime p and $n \geq 1$. Then $F^\times$ is cyclic, and its multiplicative group is equal to $\mu_{p^n-1}(F)$.*

PROOF. Since $|F^\times| = p^n - 1$, every element of $F^\times$ is a root of the polynomial $x^{p^n-1} - 1$, and conversely. Therefore, it follows from Lemma 6.5.4 that $F^\times = \mu_{p^n-1}(F)$ is cyclic. □

COROLLARY 6.5.6. *The group $(\mathbb{Z}/p\mathbb{Z})^\times$ of units in $\mathbb{Z}/p\mathbb{Z}$ is cyclic of order $p-1$.*

EXAMPLE 6.5.7. The cyclic group $(\mathbb{Z}/17\mathbb{Z})^\times$ of order 16 is generated by 3.

LEMMA 6.5.8. *Let F be a field of characteristic a prime p, and let $\alpha, \beta \in F$. Then we have*

$$(\alpha + \beta)^{p^n} = \alpha^{p^n} + \beta^{p^n}$$

for all $n \geq 0$.

PROOF. It is easy to see that $\binom{p}{i} \equiv 0 \mod p$ for $1 \le i \le p-1$, and so we have the result for $n = 1$ by the binomial theorem. By induction, the result for general n follows immediately. □

THEOREM 6.5.9. *Let n be a positive integer. There exists a field $\mathbb{F}_{p^n}$ of order p^n containing $\mathbb{F}_p$, and it is unique up to isomorphism. Moreover, if E is a finite field extension of $\mathbb{F}_p$ of degree a multiple of n, then E contains a unique subfield isomorphic to $\mathbb{F}_{p^n}$.*

PROOF. Let F be the set of roots of $x^{p^n} - x$ in a splitting field Ω of $x^{p^n} - x$ over $\mathbb{F}_p$. If $\alpha, \beta \in F$ are nonzero, then clearly $(\alpha\beta^{-1})^{p^n} = \alpha\beta^{-1}$, so $\alpha\beta^{-1} \in F$. Moreover, we have $(\alpha - \beta)^{p^n} = \alpha^{p^n} - \beta^{p^n}$ by Lemma 6.5.8, so $(\alpha - \beta)^{p^n} = \alpha - \beta$. It follows that F is a field in which $x^{p^n} - x$ splits, so it equals Ω.

Now, F has at most p^n elements by definition. We must show that has exactly p^n elements, so that its degree is n over $\mathbb{F}_p$. Clearly x factors into $x^{p^n} - x$ exactly once. Let $a \in F^\times$, and set

$$g(x) = \frac{x^{p^n} - x}{x - a} = \sum_{i=1}^{p^n - 1} a^{i-1} x^{p^n - i}.$$

Then we have

$$g(a) = \sum_{i=1}^{p^n - 1} a^{p^n - 1} = (p^n - 1)a^{p^n - 1} = -1 \neq 0,$$

so $x - a$ is not a factor of $g(x)$, and therefore all roots of $x^{p^n} - x$ are distinct.

We prove the remaining claims. First, any finite field extension of $\mathbb{F}_p$ of degree a multiple m of n has p^m elements, and Proposition 6.5.5 then implies that that it consists of roots of $x^{p^m} - x$. In particular, it contains a unique subfield of degree n consisting of the roots of $x^{p^n} - x$. Next, note that $F \cong F[x]/(f)$, where f is the minimal polynomial of a generator of $\mu_{p^n - 1}(F)$. Given any other field F' of order p^n, it also consists of the roots of $x^{p^n} - x$, so contains a root of f. This root then generates F', being a primitive $(p^n - 1)$th root of unity, so $F' \cong F[x]/(f)$ as well. □

REMARK 6.5.10. Since $\mathbb{F}_{p^n}$ has order p^n and is an $\mathbb{F}_p$-vector space, we have $[\mathbb{F}_{p^n} : \mathbb{F}_p] = n$.

COROLLARY 6.5.11. *The field $\mathbb{F}_{p^n}$ contains a subfield isomorphic to $\mathbb{F}_{p^m}$ if and only if m divides n.*

From now on, for a prime p and a positive integer n, we will speak of $\mathbb{F}_{p^n}$ as being the unique (up to isomorphism) field of order p^n.

EXAMPLE 6.5.12. The field $\mathbb{F}_9$ consists of 0 and 8th roots of unity. We have $\mathbb{F}_9 = \mathbb{F}_3(\zeta)$, where ζ is a primitive 8th root of unity (or even a primitive fourth root of unity), so a root of $x^4 + 1$. Since $[\mathbb{F}_9 : \mathbb{F}_3] = 2$, the minimal polynomial of ζ must be of degree 2. Over $\mathbb{F}_3$, we have only three irreducible polynomials of degree two: $x^2 + 1$, $x^2 + x - 1$ and $x^2 - x - 1$. The product of the latter two is $x^4 + 1$, which is to say that the 2 of the primitive 8th roots of unity have minimal polynomial $x^2 + x - 1$ and the other two $x^2 - x - 1$. On the other hand, we have $\mathbb{F}_9 = \mathbb{F}_3(\zeta^2)$ as well, and ζ^2 is a primitive 4th root of unity with minimal polynomial $x^2 + 1$.

The following result is rather useful.

PROPOSITION 6.5.13. *Let q be a power of a prime p. Let $m \geq 1$, and let ζ_m be a primitive mth root of unity in an extension of $\mathbb{F}_p$. Then $[\mathbb{F}_q(\zeta_m) : \mathbb{F}_q]$ is the order k of q in $(\mathbb{Z}/m\mathbb{Z})^\times$. In other words, we have $\mathbb{F}_q(\zeta_m) = \mathbb{F}_{q^k}$.*

PROOF. Let $k = [\mathbb{F}_q(\zeta_m) : \mathbb{F}_q]$. Then $\mathbb{F}_q(\zeta_m) = \mathbb{F}_{q^k}$, and so m divides $q^k - 1$, and then q has order dividing k in $(\mathbb{Z}/m\mathbb{Z})^\times$. On the other hand, since $\mathbb{F}_q(\zeta_m)$ is not contained in $\mathbb{F}_{q^j}$ for any $j < k$, we have that q^j is not 1 in $(\mathbb{Z}/m\mathbb{Z})^\times$. That is, q has the desired order k modulo m. □

In order to apply the previous result, it is useful to understand the structure of the unit group of $\mathbb{Z}/m\mathbb{Z}$.

PROPOSITION 6.5.14. *Let m be a positive integer, and write $m = p_1^{r_1} p_2^{r_2} \cdots p_k^{r_k}$ for distinct primes p_i and positive integers r_i for $1 \leq i \leq k$, for some $k \geq 1$. Then*

$$(\mathbb{Z}/m\mathbb{Z})^\times \cong \prod_{i=1}^{k} (\mathbb{Z}/p_i^{r_i}\mathbb{Z})^\times.$$

Moreover, if p is a prime number and r is a positive integer, we have

$$(\mathbb{Z}/p^r\mathbb{Z})^\times \cong \begin{cases} \mathbb{Z}/(p-1)\mathbb{Z} \times \mathbb{Z}/p^{r-1}\mathbb{Z} & \text{if } p \text{ is odd} \\ \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2^{r-2}\mathbb{Z} & \text{if } p = 2 \text{ and } r \geq 2. \end{cases}$$

PROOF. The first statement is a corollary of the Chinese remainder theorem for $\mathbb{Z}$. The reduction map $(\mathbb{Z}/p^r\mathbb{Z})^\times \to (\mathbb{Z}/p\mathbb{Z})^\times \cong \mathbb{Z}/(p-1)\mathbb{Z}$ (noting Corollary 6.5.6) then has kernel the multiplicative group $(1+p\mathbb{Z})/(1+p^r\mathbb{Z})$ of order p^{r-1}. If p is odd, then $(1+p)^{p^{i-1}} - 1 \equiv p^i \bmod p^{i+1}$ by the binomial theorem, so $1+p$ has order p^{r-1} in the group. If $p = 2$ and $r \geq 2$, then $5 = 1+4$ similarly generates the subgroup $(1+4\mathbb{Z})/(1+2^r\mathbb{Z})$ of order 2^{r-2}. Clearly, this group does not contain -1, which has order 2. That is, $(1+2\mathbb{Z})/(1+2^r\mathbb{Z})$ is generated by the images of -1 and 5 and so is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2^{r-2}\mathbb{Z}$. □

6.6. Cyclotomic fields

Let us explore the extensions of $\mathbb{Q}$ generated by roots of unity, known as cyclotomic fields.

NOTATION 6.6.1. Let $n \geq 1$. We will use ζ_n to denote a primitive nth root of unity in an extension of $\mathbb{Q}$. We can and therefore do choose these so that $\zeta_n^{n/m} = \zeta_m$ if m divides n. For instance, one could take $\zeta_n = e^{2\pi i/n} \in \mathbb{C}$.

DEFINITION 6.6.2. Let $n \geq 1$. Then nth *cyclotomic field* is the extension of $\mathbb{Q}$ generated by a primitive nth root of unity ζ_n.

REMARK 6.6.3. The nth cyclotomic field $\mathbb{Q}(\zeta_n)$ is Galois over $\mathbb{Q}$. That is, $\mathbb{Q}(\zeta_n)$ is the splitting field of $x^n - 1$ in that all of the roots of $x^n - 1$ are powers of ζ_n.

DEFINITION 6.6.4. The nth *cyclotomic polynomial* Φ_n is the unique monic polynomial in $\mathbb{Q}[x]$ with roots the primitive nth roots of unity.

Note that Φ_n lies in $\mathbb{Q}[x]$ since every conjugate of a primitive nth root of unity is also a root of Φ_n. In Example 5.3.4, we saw that every

$$\Phi_p = 1 + x + \cdots + x^{p-1},$$

where p is prime, is irreducible using the Eisenstein criterion.

REMARKS 6.6.5. Let n be a positive integer.

a. We have

$$x^n - 1 = \prod_{d \mid n} \Phi_d,$$

with the sum taken over positive divisors of n.

b. Every conjugate to a primitive nth root of unity is also necessarily a root of $x^n - 1$ that is not a root of its divisor $x^m - 1$ for any m dividing n, which is to say another primitive nth root of unity. Therefore Φ_n as defined lies in $\mathbb{Q}[x]$.

c. We have

$$\Phi_n(x) = \prod_{\substack{i=1 \\ \gcd(i,n)=1}}^{n} (x - \zeta_n^i),$$

and therefore Φ_n has degree $\varphi(n)$, where φ is the Euler-phi function. In particular, we have $\deg \Phi_n = \varphi(n)$.

DEFINITION 6.6.6. The *Möbius function* $\mu \colon \mathbb{Z}_{>0} \to \{-1, 0, 1\}$ is defined by

$$\mu(n) = \begin{cases} (-1)^k & \text{if } n \text{ is a product of } k \text{ distinct primes,} \\ 0 & \text{otherwise.} \end{cases}$$

We note the following.

LEMMA 6.6.7. *For any $n \geq 2$, one has $\sum_{d \mid n} \mu(d) = 0$.*

PROOF. Since $\mu(d)$ is zero if d is divisible by a square of a prime, we have $\sum_{d \mid n} \mu(d) = \sum_{d \mid m} \mu(d)$, where m is the product of the primes dividing n. If there are k such primes, then there are $\binom{k}{j}$ products of j of them, each of which contributes $(-1)^j$ to the sum. In other words,

$$\sum_{d \mid n} \mu(d) = \sum_{j=0}^{k} \binom{k}{j} (-1)^j = (1-1)^k = 0,$$

since $k \geq 1$. □

THEOREM 6.6.8 (Möbius inversion formula). *Let A be an abelian group and $f \colon \mathbb{Z}_{>0} \to A$ a function. Define $g \colon \mathbb{Z}_{>0} \to A$ by*

$$g(n) = \sum_{d \mid n} f(d)$$

for $n \geq 1$. Then

$$f(n) = \sum_{d \mid n} \mu(d) g(\tfrac{n}{d}).$$

PROOF. We calculate

$$\sum_{d|n}\mu(\tfrac{n}{d})g(d)=\sum_{d|n}\sum_{k|d}\mu(\tfrac{n}{d})f(k)=\sum_{k|n}\sum_{\substack{d|n\\k|d}}\mu(\tfrac{n}{d})f(k)=\sum_{k|n}\sum_{c|\frac{n}{k}}\mu(\tfrac{n}{kc})f(k)=f(n),$$

the last step by Lemma 6.6.7. □

Taking $A=\mathbb{Q}(x)^{\times}$, we have the following.

LEMMA 6.6.9. *Let $n\geq 1$. Then*

$$\Phi_n=\prod_{\substack{d|n\\d\geq 1}}(X^{n/d}-1)^{\mu(d)}.$$

The lemma can be used to calculate cyclotomic polynomials explicitly.

EXAMPLES 6.6.10.

a. We have $\Phi_1(x)=x-1$.

b. For a prime p and $k\geq 1$, we have

$$\Phi_{p^k}(x)=\frac{x^{p^k}-1}{x^{p^{k-1}}-1}=\sum_{i=0}^{p-1}x^{ip^{k-1}}.$$

c. For p and q distinct primes, we have

$$\Phi_{pq}(x)=\frac{(x^{pq}-1)(x-1)}{(x^p-1)(x^q-1)}=\frac{\Phi_q(x^p)}{\Phi_q(x)}$$

For instance, taking $q=2$ we obtain

$$\Phi_{2p}(x)=\frac{x^p+1}{x+1}=\Phi_p(-x),$$

and we have

$$\Phi_{15}(x)=x^8-x^7+x^5-x^4+x^3-x+1.$$

The nth cyclotomic polynomial is in fact irreducible over $\mathbb{Q}$.

THEOREM 6.6.11. *Let $n\geq 1$. Then the cyclotomic polynomial Φ_n is irreducible in $\mathbb{Q}[x]$.*

PROOF. Write $\Phi_n=fg$ with $f,g\in\mathbb{Z}[x]$ and f monic irreducible with ζ as a root. Take any prime p not dividing n, and note that ζ^p is also a root of Φ_n.

If ζ^p is a root of g, then $g(x^p)$ is divisible by the minimal polynomial $f(x)$ of ζ. Let $\bar{f}$ and $\bar{g}$ denote the reductions modulo p of f and g respectively. Then $\bar{g}(x^p)=\bar{g}(x)^p\in\mathbb{F}_p[x]$ is divisible by $\bar{f}(x)$, so $\bar{g}$ and $\bar{f}$ have a common factor. The reduction $\phi_n=\bar{f}\bar{g}$ of Φ_n modulo p therefore has a multiple root in $\overline{\mathbb{F}_p}$. In particular, x^n-1 has a multiple root, but we know that it does not. That is, if we choose $k\geq 1$ so that $p^k\equiv 1 \bmod n$, then the cyclic group $\mathbb{F}_{p^k}^{\times}$ of order p^k-1 contains n distinct nth roots of unity.

Thus, ζ^p is a root of f for any prime p and any root ζ of f. Since any integer a prime to n can be written as a product of primes not dividing n, it follows that ζ^a is a root of f for all a prime to p. This forces $f=\Phi_n$, so Φ_n is irreducible. □

6.7. Field embeddings

DEFINITION 6.7.1. Let E and E' be extensions of a field F, and let $\varphi\colon E \to E'$ be an isomorphism of fields. We say that φ *fixes* F if $\varphi(\alpha) = \alpha$ for all $\alpha \in F$.

DEFINITION 6.7.2. Let α and β be elements of field extensions of a field F. We say that α and β are *conjugate* over F if there exists a field isomorphism $\varphi\colon F(\alpha) \to F(\beta)$ fixing F such that $\varphi(\alpha) = \beta$.

PROPOSITION 6.7.3. *Let E and E' be extensions of a field F, and let $\alpha \in E$, $\beta \in E'$ be algebraic over F. Then α and β are conjugate over F if and only if the minimal polynomials of α and β in $F[x]$ are equal.*

PROOF. Suppose that α and β are conjugate over F, and let $\varphi\colon F(\alpha) \to F(\beta)$ be a field isomorphism such that $\varphi(\alpha) = \beta$ and φ restricts to the identity map on F. Then $\varphi(g(\alpha)) = g(\beta)$ for all $g \in F[x]$. Let $f \in F[x]$ be the minimal polynomial of α. Then we have

$$0 = \varphi(0) = \varphi(f(\alpha)) = f(\beta),$$

so β is a root of f. As f is irreducible, it must be the minimal polynomial of β.

Conversely, suppose that α and β have the same minimal polynomial $f \in F[x]$. Then we have isomorphisms from $F[x]/(f)$ to $F(\alpha)$ and $F(\beta)$ as in Theorem 6.1.15, and composing the inverse of the first with the latter yields the desired isomorphism $F(\alpha) \to F(\beta)$. □

EXAMPLE 6.7.4. Since i and $-i$ are both roots of the irreducible polynomial x^2+1 over $\mathbb{R}$, they are conjugate elements of $\mathbb{C}$. Therefore, there is a field isomorphism $\mathbb{C} \to \mathbb{C}$ that takes i to $-i$ and fixes $\mathbb{R}$. Such an isomorphism must take $a+bi$ to its complex conjugate

$$\overline{a+bi} = a-bi$$

and is therefore the usual complex conjugation. In particular, complex conjugation is an isomorphism of fields, which is also easily verified directly. Moreover, we see that if $f \in \mathbb{R}[x]$ has a root α, then $\bar{\alpha}$ is a root as well, since $\overline{f(\alpha)} = f(\bar{\alpha})$.

DEFINITION 6.7.5. An *embedding of fields*, or *field embedding*, is an injective ring homomorphism $\varphi\colon F \to F'$, where F and F' are fields.

REMARK 6.7.6. Any nonzero ring homomorphism between fields is injective, so "injective" can be replaced by "nonzero" in the definition of a field embedding.

DEFINITION 6.7.7. Let $\varphi\colon F \to M$ be a field embedding, and let E/F be an extension field. We say that a field embedding $\Phi\colon E \to M$ *extends* φ, and is an *extension* of φ, if $\Phi|_E = \varphi$.

EXAMPLE 6.7.8. We have a field embedding $\iota\colon \mathbb{Q} \to \mathbb{R}$. There are two field embeddings $\iota'\colon \mathbb{Q}(\sqrt{2}) \to \mathbb{R}$ extending ι. Either we take $\iota'(a+b\sqrt{2}) = a+b\sqrt{2}$ for $a,b \in \mathbb{Q}$, or we set $\iota'(a+b\sqrt{2}) = a-b\sqrt{2}$. On the other hand, there is no field embedding $\kappa\colon \mathbb{Q}(i) \to \mathbb{R}$ extending ι, since there is no element of i that would satisfy $\kappa(i)^2+1 = 0$, but there is no element of $\mathbb{R}$ with this property.

Let us give a slight extension of one direction of Proposition 6.7.3.

THEOREM 6.7.9. *Let E/F be a field extension, and let $\alpha \in E$ be algebraic over F. Let $\varphi\colon F \to M$ be a field embedding, and consider the induced map $\tilde{\varphi}\colon F[x] \to M[x]$. Let $f \in F[x]$ be the minimal polynomial of α. Then there is a bijection between the set of field embeddings $F(\alpha) \to M$ extending φ and the set of roots of $\tilde{\varphi}(f)$ in M taking an extension Φ of φ to $\Phi(\alpha)$.*

PROOF. Suppose that β is a root of $\tilde{\varphi}(f)$. Let $e_\beta\colon M[x] \to M$ denote the evaluation map at β. The composition $e_\beta \circ \tilde{\varphi}$ has kernel containing (f), and the kernel then equals (f) by the maximality of (f) and the fact that the composition is nonzero. The first isomorphism theorem yields a field embedding $F[x]/(f) \to M$ sending the coset of x to β. The map Φ is then obtained by composing with the isomorphism $F(\alpha) \to F[x]/(f)$ of Theorem 6.1.15, and it sends α to β. Moreover, if κ is any other lift of φ such that $\kappa(\alpha) = \beta$, we have

$$\kappa\left(\sum_{i=0}^{\deg f-1} c_i\alpha^i\right) = \sum_{i=0}^{\deg f-1} \varphi(c_i)\beta^i = \Phi\left(\sum_{i=0}^{\deg f-1} c_i\alpha^i\right)$$

for all $c_i \in F$ for $1 \le i < \deg f$, so $\kappa = \Phi$.

Conversely, suppose $\Phi\colon F(\alpha) \to M$ is an extension of φ. Then we have $\tilde{\varphi}(f)(\Phi(\alpha)) = \Phi(f(\alpha)) = 0$. □

COROLLARY 6.7.10. *Let E/F be a field extension, let $\alpha \in E$ be algebraic over F, and let $\varphi\colon F \to M$ be a field embedding. Let $\tilde{\varphi}\colon F[x] \to M[x]$ denote the induced map on polynomial rings. The number of extensions of φ to an embedding $\Phi\colon F(\alpha) \to M$ is the number of distinct roots of $\tilde{\varphi}(f)$ in M, where $f \in F[x]$ is the minimal polynomial of α.*

REMARK 6.7.11. In the setting of Corollary 6.7.10, we may identify F with its isomorphic image $\varphi(F)$. This allows us to think of F as a subfield of M. In this case, $f \in F[x]$ may be thought of as itself having roots in M, and the number of embeddings of $F(\alpha)$ in M is the number of distinct roots of f in M.

In general, for finite extensions, we have the following.

COROLLARY 6.7.12. *Let E/F be a finite extension of fields. Let $\varphi\colon F \to M$ be a field embedding. Then the number of extensions $\Phi\colon E \to M$ of F is finite, less than or equal to $[E : F]$.*

PROOF. Since any finite extension E/F is finitely generated, it suffices by the multiplicativity of degrees of field extensions in Corollary 6.2.12 and recursion to prove the result in the case that $E = F(\alpha)$ for some $\alpha \in E$. In this case, the degree of the minimal polynomial of α is equal to $[E : F]$ and is greater than or equal to the number of distinct roots in M of the image of the minimal polynomial of α. The result is therefore a consequence of Corollary 6.7.10. □

EXAMPLE 6.7.13. As seen in Example 6.7.8, there are exactly two embeddings of $\mathbb{Q}(\sqrt{2})$ in $\mathbb{R}$, but no embeddings of $\mathbb{Q}(i)$ in $\mathbb{R}$.

EXAMPLE 6.7.14. There are four embeddings of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ in $\mathbb{R}$. If φ is such an embedding, then we have $\varphi(\sqrt{2}) = \pm\sqrt{2}$ and $\varphi(\sqrt{3}) = \pm\sqrt{3}$, and the signs determine the embedding uniquely.

Finally, we note the following.

PROPOSITION 6.7.15. *Let E/F be an algebraic field extension, and let $\sigma\colon E \to E$ be a field embedding fixing F. Then σ is an isomorphism.*

PROOF. Let $\beta \in E$, and let $f \in F[x]$ be its minimal polynomial. By Proposition 6.7.3, every root of f in E is sent by σ to another root of f in E. As σ is injective and the set of roots of f in E is finite, σ permutes these roots. In particular, there exists a root α of f in E such that $\sigma(\alpha) = \beta$. Therefore, we have $\sigma(E) = E$, as desired. □

6.8. Algebraically closed fields

We begin with the notion of an algebraically closed field.

DEFINITION 6.8.1. A field L is *algebraically closed* if contains a root of every nonconstant polynomial $f \in L[x]$.

The following theorem has analytic, topological, geometric, and algebraic proofs (though all in a sense require some very basic analysis).

THEOREM 6.8.2 (Fundamental theorem of algebra). *The field $\mathbb{C}$ of complex numbers is algebraically closed.*

We defer an algebraic proof of this theorem until after our treatment of Galois theory. For the reader's enjoyment, here are sketches of three proofs which require some knowledge of subjects outside of this course. The first two use complex analysis:

REMARK 6.8.3. if $p \in \mathbb{C}[x]$ has no roots, then p^{-1} is homolorphic and bounded as a function on $\mathbb{C}$, hence constant by the maximum modulus principle.

REMARK 6.8.4. A nonconstant polynomial $p \in \mathbb{C}[x]$ defines a nonconstant continuous map from the Riemann sphere $\mathbb{P}^1(\mathbb{C})$ to itself. Its image is closed as $\mathbb{P}^1(\mathbb{C})$ is compact Hausdorff, while its image is open by the holomorphicity of p and the open mapping theorem, so the image is $\mathbb{P}^1(\mathbb{C})$.

Next, algebraic topology:

REMARK 6.8.5. Suppose that $p \in \mathbb{C}[x]$ is monic of degree n, and choose $r > 0$ such that $|p(z) - z^n| < r^n$ for all $z \in \mathbb{C}$ with $|z| = r$. The map $F\colon S^1 \to S^1$ with

$$F(z) = \frac{p(rz)}{|p(rz)|}$$

is homotopic to $z \mapsto z^n$ by a homotopy $H\colon [0,1] \times S^1 \to S^1$ given by

$$H(t,z) = \frac{tp(rz) + (1-t)(rz)^n}{|tp(rz) + (1-t)(rz)^n|}.$$

Now F extends to a map $\mathbb{C} \to S^1$ on the simply connected space $\mathbb{C}$ by the same formula, so induces the zero map on $\pi_1(S^1) \cong \mathbb{Z}$. But $z \mapsto z^n$ induces multiplication by n on $\pi_1(S^1)$, so $n = 0$.

PROPOSITION 6.8.6. *Let L be an algebraically closed field, and let $f \in L[x]$ be nonconstant. Then f splits in L.*

PROOF. We prove this by induction, as it is clear for $\deg f = 1$. Suppose we know the result for all polynomials of degree less than $n = \deg f$. Since f has a root α in L, we have $f = (x-\alpha)g$ for some $g \in L[x]$ of degree $n-1$. By induction, g factors into linear terms. □

COROLLARY 6.8.7. *Let M be an algebraic extension of an algebraically closed field L. Then $M = L$.*

PROOF. Let $\alpha \in M$. As M is algebraic over L, there exists a nonconstant $f \in L[x]$ with $f(\alpha) = 0$, and by Proposition 6.8.6, the polynomial f is divisible by $x-\alpha$ in $L[x]$ (recalling that $M[x]$ is a UFD). Therefore, we have $\alpha \in L$. □

We next show that extensions of field embeddings into algebraically closed fields always exist, when the extension is algebraic.

THEOREM 6.8.8. *Let E/F be an algebraic extension of fields. Let $\varphi \colon F \to M$ be a field embedding, where M is an algebraically closed field. Then there exists a field embedding $\Phi \colon E \to M$ extending φ.*

PROOF. Let X denote the nonempty set of all pairs (K,σ), where K is an intermediate subfield of E/F and $\sigma \colon K \to M$ is an extension of φ. We say that $(K,\sigma) \le (K',\sigma')$ for (K,σ) and $(K',\sigma') \in X$ if K' contains K and $\sigma'|_K = \sigma$. Let C be a chain in X, set

$$L = \bigcup_{(K,\sigma)\in C} K$$

and define $\tau \colon L \to M$ by $\tau|_K = \sigma$ for all $(K,\sigma) \in C$. It is easy to see that τ is a well-defined field embedding, since C is a chain, and therefore, $(L,\tau) \in X$ is an upper bound for C.

By Zorn's lemma, we therefore have that X contains a maximal element, which we call (Ω,λ). We claim that $E = \Omega$. To see this, let $\alpha \in E$, and let $f \in \Omega[x]$ be the minimal polynomial of α over Ω. If $f = \sum_{i=0}^n a_i x^i$ with $a_i \in \Omega$ for $0 \le i \le n$ and $n = \deg f$, then we set

$$g = \sum_{i=0}^n \lambda(a_i) x^i.$$

Since M is algebraically closed, g has a root β in M. By Proposition 6.7.9, we may then extend λ to an embedding $\lambda' \colon \Omega(\alpha) \to M$. We then have $(\Omega,\lambda) \le (\Omega(\alpha),\lambda')$, and the maximality of Ω forces $E = \Omega$. Setting $\Phi = \lambda$, we are done. □

PROPOSITION 6.8.9. *The set of all algebraic elements over a field F in an extension E is a subfield of E, and it is equal to the the largest intermediate extension of E/F that is algebraic over F.*

PROOF. Let M denote the set of all algebraic elements over F in E, and let $\alpha,\beta \in M$. Then $F(\alpha,\beta)/F$ is a finite extension, so every element of it is algebraic. In particular, $\alpha-\beta$ and, if $\beta \ne 0$, the element $\alpha\beta^{-1}$ are elements of $F(\alpha,\beta)$, so they are algebraic elements over F, hence contained in M. Therefore, M is a field. The second statement is then an immediate consequence of the definition of M. □

COROLLARY 6.8.10. *The set $\overline{\mathbb{Q}}$ of algebraic numbers in $\mathbb{C}$ forms a field.*

DEFINITION 6.8.11. An *algebraic closure* of a field F is an algebraically closed, algebraic extension of F.

REMARK 6.8.12. Since an algebraic closure is algebraic, every element of the the algebraic closure of a field F has to be the root of a polynomial with F-coefficients. On the other hand, since $\overline{F}$ is algebraically closed, it contains all roots of every polynomial with coefficients in $\overline{F}$ (i.e., every polynomial in $\overline{F}[x]$ factors completely). Thus, if an algebraic closure exists, and we shall see that it does, it consists exactly of all roots of polynomials in F, and every root of a polynomial with coefficients in $\overline{F}$ is actually the root of a polynomial with coefficients in F.

In fact, if a field is contained in an algebraically closed field, then we can see that it does in fact have an algebraic closure quite directly.

PROPOSITION 6.8.13. *Let F be a field, and suppose that M is an algebraically closed extension field of F. Then M contains a unique algebraic closure of F, equal to the field of elements of M that are algebraic over F.*

PROOF. Let $\overline{F}$ denote the field consisting of all elements of M that are algebraic over F. We claim that $\overline{F}$ is algebraically closed. For this, suppose that $f \in \overline{F}[x]$, and $\alpha \in M$ be a root. As α is algebraic over $\overline{F}$, we have by Proposition 6.2.19 that α is also algebraic over F. That is, α is an element of $\overline{F}$. □

COROLLARY 6.8.14. *The field $\overline{\mathbb{Q}}$ of algebraic numbers in $\mathbb{C}$ is an algebraic closure of $\mathbb{Q}$.*

Using Zorn's lemma, we may prove that every field has an algebraic closure. This is the first result on extension fields in which we do not have a previously given field that contains the field of interest, which makes the proof rather more tricky.

THEOREM 6.8.15. *Every field F has an algebraic closure.*

PROOF. Let F be a field. Let Ω be a set that is the disjoint union of finite sets R_f for each monic irreducible $f \in F[x]$, where the number of elements in R_f is the number of distinct roots of f in a splitting field. (We will end up identifying the elements of R_f with roots of f, but they do not start as such.) We may view F as a subset of Ω by identifying $a \in F$ with the unique element of R_{x-a}. Let X be the nonempty set of all algebraic extensions of F, the underlying sets of which are contained in Ω in the sense if $E \in X$, then every $\alpha \in E$ lies in R_f for $f \in F[x]$ the minimal polynomial of α. We put a partial ordering on X by $E \le E'$ for $E, E' \in X$ if and only if $E \subseteq E'$ and E'/E is a field extension.

Let $\mathscr{C}$ be chain in X, and let K be the union of the fields in $\mathscr{C}$. Then K is a field, as any two elements $\alpha, \beta \in K$ satisfy $\alpha, \beta \in E$ for some $E \in \mathscr{C}$ (taking the larger of the two fields in which α and β are contained by definition), and then $\alpha - \beta \in E$ and $\alpha\beta^{-1} \in E$ if $\beta \neq 0$. Since $K \in X$, the chain $\mathscr{C}$ has an upper bound, and we may apply Zorn's lemma to the set X to find a maximal element $\overline{F} \in X$.

Let $f \in F[x]$, and let $g \in \overline{F}[x]$ be a monic, nonconstant irreducible polynomial dividing f. We then have that $E = \overline{F}[x]/(g)$ is an extension of $\overline{F}$ that is algebraic over F. We may view the underlying set of E as being contained in Ω as follows. If $h \in F[x]$ is a monic irreducible polynomial with a root in E, we may identify those of its distinct roots in E that are not contained in $\overline{F}$ with distinct elements of R_h that are not in $\overline{F}$. Since $\overline{F} \in X$ is maximal, we must have $E = \overline{F}$. In particular, f must factor completely in $\overline{F}[x]$.

It remains only to show that $\overline{F}$ is algebraically closed. Any root β of an irreducible polynomial $g \in \overline{F}[x]$ in an extension of $\overline{F}$ is algebraic over F by Proposition 6.2.19. Therefore, g divides the minimal polynomial $f \in F[x]$ of β, which by the argument we have just given splits over $\overline{F}$. In particular, we have $\beta \in \overline{F}$. □

We next remark that the algebraic closure of any field is in fact unique up to isomorphism.

PROPOSITION 6.8.16. *Let M and M' be algebraic closures of a field F. Then there exists an isomorphism $\Phi\colon M \to M'$ fixing F.*

PROOF. Theorem 6.8.8 applied to the case that $\varphi = \mathrm{id}_F$, $E = M$, and $M = M'$ implies that there exists a field embedding $\Phi\colon M \to M'$ extending F. To see that it is an isomorphism, note that the image of Φ is algebraic over F, being contained in M', and algebraically closed over F since a root of a polynomial in $F[x]$ in M maps under Φ to a root of the same polynomial. Therefore, $\Phi(M)$ is an algebraic closure of F contained in M', and hence must be M' itself. □

REMARK 6.8.17. As any two algebraic closures of a field F are isomorphic via an isomorphism that fixes F, we usually refer to "the" algebraic closure of F, denoting it by $\overline{F}$.

REMARK 6.8.18. If E is an algebraic extension of F and $\overline{E}$ is the algebraic closure of E, then it is also an algebraic closure of F. In particular, there exists an algebraic closure of F containing E.

6.9. Transcendental extensions

DEFINITION 6.9.1. A field extension E/F is *totally transcendental* if every element of $E - F$ is transcendental over E.

TERMINOLOGY 6.9.2. For a ring R and an indexing set I, we may speak of the polynomial ring $R[(x_i)_{i\in I}]$ in the variables x_i for $i \in I$. It is simply the union over all finite lists $i_1, \dots, i_n$ of distinct elements of I of the polynomial rings $R[x_{i_1}, \dots, x_{i_n}]$, with the operations being induced by the operations on these rings. If R is commutative, then the rational function field $R((x_i)_{i\in I})$ is the fraction field of $R[(x_i)_{i\in I}]$. This field is itself the union of the rational function fields $R(x_{i_1}, \dots, x_{i_n})$.

PROPOSITION 6.9.3. *For any indexing set I, the field $F((t_i)_{i\in I})$ of rational functions in the variables t_i for $i \in I$ is purely transcendental over F.*

PROOF. Consider first the extension $F(t)/F$ given by the F-rational function field in a single variable t. Let $\alpha = \frac{f}{g} \in F(t) - F$, where $f, g \in F[t]$ and $g \neq 0$. We may view $f(x)$ and $g(x)$ as elements of $F[x]$. Then $\alpha \cdot g(x) \in F(\alpha)[x]$ is not an element of $F[x]$, so the polynomial $f(x) - \alpha g(x)$ is nonzero but does have a root t, which is then algebraic over $F(\alpha)$. Since t is transcendental over F, this forces α to be as well. This gives the result for a single variable, and the case of finitely many variables follows immediately by induction. Since $F((t_i)_{i\in I})$ is the union of the rational function fields $F(t_{i_1}, \dots, t_{i_n})$ with $i_1, \dots, i_n \in I$, the case of finitely many variables yields the general case. □

PROPOSITION 6.9.4. *Every extension of fields is a totally transcendental extension of an algebraic extension.*

PROOF. Given a field extension K/F, we may consider its subfield E of elements algebraic over F. If $\alpha \in K - F$, then α cannot be algebraic over E. That is, if it were, then it would also be algebraic over F in that E/F is algebraic. □

DEFINITION 6.9.5. Let K/F be a field extension.

a. We say that a collection $(\alpha_i)_{i\in I}$ of elements of K, is *algebraicaly independent* over F, if $f(\alpha_{i_1},\dots,\alpha_{i_n}) \neq 0$ for all nonzero polynomials $f \in F[x_1,\dots,x_n]$ and distinct elements $i_1,\dots,i_n$ for some $n \geq 1$.

b. We say that a subset S of K is *algebraically independent* over F, or *F-algebraically independent*, if $f(s_1,\dots,s_n) \neq 0$ for all nonzero polynomials f in n variables over F and distinct $s_1,\dots,s_n \in S$ for some $n \geq 1$.

Here are a couple of straightfoward lemmas.

LEMMA 6.9.6. *Let K/F be a field extension, and let $S \subset K$ be algebraically independent over F. Then $t \in K$ is transcendental over the field $F(S)$ generated by S over F if and only if $S \cup \{t\}$ is algebraically independent over F.*

LEMMA 6.9.7. . *A subset S of a field extension K of F is algebraically independent over F if and only if each $s \in S$ is transcendental over $K(S - \{s\})$.*

DEFINITION 6.9.8. A subset S of an extension K of a field F is a *transcendence basis* of K/F if and only if S is algebraically independent over F and K is algebraic over $F(S)$.

The following equivalent conditions for being a transcendence basis nearly mimic the usual equivalent conditions for a subset of a vector space to be a basis. (That is, a subset is a basis if and only if it is a maximal linearly independent subset and if and only if it is a minimal spanning set.)

PROPOSITION 6.9.9. *Let S be a subset of an extension K of a field F. The following are equivalent:*

i. S is a transcendence basis of K/F,

ii. S is a maximal F-algebraically independent subset of K,

iii. S is a minimal subset of K such that K is algebraic over $F(S)$.

PROOF. The equivalence of (i) and (ii) is a direct consequence of Lemma 6.9.6, and the equivalence of (i) and (iii) is a direct consquence of Lemma 6.9.7. □

THEOREM 6.9.10. *Every F-algebraically independent subset of an extension K/F is contained in a transcendence basis, and every subset of K that generates an extension over which K is algebraic contains a transcendence basis.*

PROOF. Let A be an F-algebraically independent subset of K. Let X be the set of F-algebraically independent subsets of K containing A, ordered by inclusion. We may take the union of any chain $\mathscr{C}$ in X, and it is F-algebraically independent in that any finitely many elements of the union on which we would test algebraic independence is contained in some element of the chain. This union is an upper bound, and thus by Zorn's lemma, X contains a maximal element B. To finish the proof, we need only see that K is algebraic over $F(B)$. But this is clear, since if $t \in K - B$ is transcendental over $F(B)$, then $B \cup \{t\}$ is F-algebraically independent, contradicting the maximality of B.

Now, let $S \subseteq K$ be such that $K/F(S)$ is algebraic. Consider the set Y of F-algebraically independent subsets of K contained in S, again ordered by inclusion. Every chain has an upper bound as before, so Y contains a maximal element T. We need only see that K is algebraic over $F(T)$. If not, then since K is algebraic over $F(S)$ and $T \subset S$, we must have that there exists $s \in S - T$ that is transcendental over $F(T)$, and then $T \cap \{s\} \in Y$, contradicting the maximality of T. □

COROLLARY 6.9.11. *Every extension of fields K/F has an intermediate field E such that K/E is algebraic and E/F is totally transcendental.*

We omit a proof of the following.

THEOREM 6.9.12. *If S and T are transcendence bases of an extension K/F, then S and T have the same cardinality.*

In particular, we may make the following definition.

DEFINITION 6.9.13. We say that a field extension K/F has *finite transcendence degree* if it has a finite transcendence basis, in which case the number of elements in a transcendence basis is called the *transcendence degree*. Otherwise, we say that K/F has *infinite transcendence degree*.

6.10. Separable extensions

DEFINITION 6.10.1. Let F be a field. Let $f \in F[x]$ be nonzero, and let $\alpha \in \overline{F}$ be a root of f. The *multiplicity* of α as a root of f is the largest positive integer m such that $(x-\alpha)^m$ divides f in $\overline{F}[x]$.

EXAMPLE 6.10.2. Let $f = x^p - t \in \mathbb{F}_p(t)[x]$, which is irreducible. In $\overline{\mathbb{F}_p(t)}[x]$, we have

$$f = x^p - t = (x - t^{1/p})^p,$$

so $t^{1/p}$ has multiplicity p as a root of f.

LEMMA 6.10.3. *Let F be a field, and let $f \in F[x]$ be irreducible. Then every root of f in an algebraic closure $\overline{F}$ of F has the same multiplicity.*

PROOF. Let $\alpha, \beta \in \overline{F}$ be roots of f. Fix an field isomorphism $\sigma\colon F(\alpha) \to F(\beta)$ taking α to β, and extend it to an embedding $\tau\colon \overline{F} \to \overline{F}$. Let $\tilde{\tau}\colon \overline{F}[x] \to \overline{F}[x]$ map induced by τ. If m denotes the multiplicity of α, then

$$\tilde{\tau}((x-\alpha)^m) = (x-\beta)^m.$$

Since $(x-\alpha)^m$ divides f in $\overline{F}[x]$ and $\tilde{\tau}(f) = f$, the multiplicity of β is then at least m, but this was independent of the choice of α and β, so α and β have the same multiplicity. □

COROLLARY 6.10.4. *Let F be a field. The number of distinct roots of an irreducible polynomial $f \in F[x]$ in an algebraic closure $\overline{F}$ of F divides the degree of f.*

DEFINITION 6.10.5. Let F be a field. We say that a nonconstant polynomial $f \in F[x]$ is *separable* if every root of f has multiplicity 1.

DEFINITION 6.10.6. Let F be a field and $\overline{F}$ be an algebraic closure of F. An element $\alpha \in \overline{F}$ is *separable* over F if and only if its minimal polynomial is separable over F.

DEFINITION 6.10.7. We say that an algebraic extension E/F is *separable* if every $\alpha \in E$ is separable over F.

NOTATION 6.10.8. Let K and L be extensions of a field F. We will denote the set of field embeddings of K into L that fix F by $\mathrm{Emb}_F(K,L)$. If K is algebraic over F and L is taken to be a fixed algebraic closure of F, we will simply write $\mathrm{Emb}_F(K)$ (despite the dependence on the algebraic closure).

LEMMA 6.10.9. *Let E/F be a field extension, and let $\alpha \in E$ be algebraic over F. Then α is separable over F if and only if $F(\alpha)/F$ is separable.*

PROOF. We prove the nontrivial direction, which results from several applications of Theorem 6.7.9. Fix an algebraic closure $\overline{F}$ of F. For a given $\beta \in F(\alpha)$, the number $e = |\mathrm{Emb}_F(F(\beta))|$ is at most the degree $[F(\beta) : F]$, with equality if and only if β is separable. Since α is separable over F, we have

$$|\mathrm{Emb}_F(F(\alpha))| = [F(\alpha) : F].$$

Moreover, α is separable over $F(\beta)$ as well, since its minimal polynomial over $F(\beta)$ divides its minimal polynomial over F. Thus, the number of embeddings of $F(\alpha)$ in $\overline{F}$ extending a given embedding of $F(\beta)$ into $\overline{F}$ is exactly $[F(\alpha) : F(\beta)]$. Therefore, we have that

$$[F(\alpha) : F] = [F(\alpha) : F(\beta)]e,$$

which means that $e = [F(\beta) : F]$, so β is separable. □

LEMMA 6.10.10. *Let E be an algebraic extension of a field F, and let K be an algebraic extension of E. If K/F is separable, then so are K/E and E/F.*

PROOF. Suppose that K/F is separable. By definition, if $\alpha \in E$, then $\alpha \in K$, so its minimal polynomial over F is separable. Moreover, the minimal polynomial of any $\beta \in K$ over E divides the minimal polynomial of β over F, so β is separable over E. □

We also have the following.

PROPOSITION 6.10.11. *Let E/F be a finite extension. Fix an algebraic closure $\overline{F}$ of F.*

a. The number of embeddings of E into $\overline{F}$ that fix F divides $[E : F]$.

b. The number of embeddings of E into $\overline{F}$ that fix F is equal to $[E : F]$ if and only if E/F is separable.

PROOF. Let e denote the number of embeddings of E in $\overline{F}$. Write $E = F(\alpha_1, \alpha_2, \dots, \alpha_n)$, and let $E_i = F(\alpha_1, \alpha_2, \dots, \alpha_{i-1})$ for $1 \le i \le n+1$. Then $E_{i+1} = E_i(\alpha_i)$ for $i \le n$, and by Theorem 6.7.9, the number e_i of embeddings of E_{i+1} into $\overline{F}$ extending an embedding φ_i of E_i into $\overline{F}$ is the number of distinct roots of the minimal polynomial of α_i over E_i. This number, in turn, is a divisor of $[E_{i+1} : E_i]$, with equality if and only if α_i is separable over E_i. Since

$$e = \prod_{i=1}^{n} e_i$$

and

$$[E : F] = \prod_{i=1}^{n} [E_{i+1} : E_i],$$

we therefore have that e divides $[E : F]$, with equality if and only if $e_i = [E_{i+1} : E_i]$ for each i, and in particular, noting Lemma 6.10.10, if E/F is separable.

Conversely, suppose that $e = [E : F]$. For $\beta \in E$, the number of distinct roots c of its minimal polynomial is the number of embeddings of $F(\beta)$ into $\overline{F}$ fixing F. By the above argument, the number of embeddings d of E into $\overline{F}$ extending one of those embeddings divides $[E : F(\beta)]$, and we have $e = cd$, so we must have $c = [F(\beta) : F]$. That is, E/F is separable. □

PROPOSITION 6.10.12. *Let K be an algebraic extension of a field F, and let E be an intermediate field in K/F. Then K/F is separable if and only if K/E and E/F are.*

PROOF. By Lemma 6.10.10, we are reduced to showing that if K/E and E/F are separable, then K/F is separable. Proposition 6.10.11 implies this immediately if K/F is finite. In general, take $\alpha \in K$, and note that any minimal polynomial g of α over E actually has coefficients in some finite subextension E' of E, in that E/F is algebraic. Then $E'(\alpha)/E'$ is separable since g is, and E'/F is separable by Lemma 6.10.10. As $E'(\alpha)/F$ is finite, we have the result. □

DEFINITION 6.10.13. We say an extension E/F is *purely inseparable* if E contains no nontrivial separable subextensions of F.

Proposition 6.10.12 tells us that it suffices to check the separability of an extension on a generating set. It also implies the following.

COROLLARY 6.10.14. *Let K/F be an algebraic extension. The set E of all separable elements in K/F is a subfield of K. Moreover, the extension K/E is purely inseparable.*

DEFINITION 6.10.15. Let K/F be a finite extension, and let E be the maximal separable subextension of F in K.

i. The *degree of separability* $[K : F]_{\mathrm{s}}$ of K/F is $[E : F]$.

ii. The *degree of inseparability* $[K : F]_{\mathrm{i}}$ of K/F is $[K : E]$.

Finally, let us investigate circumstances under which all finite extensions of a given field are separable.

DEFINITION 6.10.16. A field F is *perfect* is every finite extension of it is separable.

EXAMPLE 6.10.17. The field $\mathbb{F}_p$ is perfect. To see this, recall the field $\mathbb{F}_{p^n}$ for $n \geq 1$ is equal to the set of roots of the polynomial $x^{p^n} - x$, which are all distinct (since there need to be p^n of them). Since the minimal polynomial of any $\alpha \in \mathbb{F}_{p^n}$, divides $x^{p^n} - x$, that polynomial is separable, and therefore $\mathbb{F}_{p^n}/\mathbb{F}_p$ is separable.

LEMMA 6.10.18. *Let E/F be an algebraic field extension. Let $f \in E[x]$ be monic, and let $m \geq 1$ be such that $f^m \in F[x]$. Then, either $m = 0$ in F or $f \in F[x]$.*

PROOF. Suppose that $f \notin F[x]$. Write $f = \sum_{i=0}^{n} a_i x^i$ with $n = \deg f$ and $a_n = 1$. Let $i \leq n-1$ be maximal such that $a_i \notin F$. The coefficient c of $x^{(m-1)n+i}$ in f^m is a polynomial in the coefficients $a_i, a_{i+1}, \dots, a_{n-1}$ such that $c - ma_i$ is a polynomial in $a_{i+1}, \dots, a_{n-1}$, which are elements of F. Since $c \in F$, we have $ma_i \in F$, which forces either $m = 0$ in F or $a_i \in F$. □

THEOREM 6.10.19. *Let F be a field of characteristic* 0*. Then F is perfect.*

PROOF. If $f \in F[x]$ is irreducible, then every root of f in an algebraic closure $\overline{F}$ occurs with some multiplicity $m \geq 1$. It follows that

$$f = \prod_{i=1}^{d} (x - \alpha_i)^m$$

for some $d \geq 1$ and distinct $\alpha_1, \alpha_2, \dots, \alpha_m \in \overline{F}$, so $f = g^m$ for some $g \in \overline{F}[x]$. Since the characteristic of F is zero, Lemma 6.10.18 tells us that $m = 1$. □

The following tells us that the degree of inseparability of a finite field extension is the power of the characteristic of the fields.

PROPOSITION 6.10.20. *Let F be a field of characteristic p. If E/F is purely inseparable and $\alpha \in E$, then $\alpha^{p^k} \in F$ for some minimal $k \geq 0$, and the minimal polynomial of α over F is $x^{p^k} - \alpha^{p^k} = (x - \alpha)^{p^k}$.*

PROOF. Fix an algebraic closure $\overline{F}$ of F containing E. Let $f \in F[x]$ be the minimal polynomial of some element α of E not in F. Again we have

$$f = \prod_{i=1}^{d} (x - \alpha_i)^m$$

for some $d \geq 1$ and distinct $\alpha_1, \alpha_2, \dots, \alpha_m \in \overline{F}$, so $f = g^m$ for some $g \in \overline{F}[x]$. We must show that m is a p-power and $d = 1$.

Write $m = p^k t$ with $p \nmid t$ and $k \geq 1$. The fact that $f = (g^{p^k})^t \in F[x]$ forces $g^{p^k} \in F[x]$ by Lemma 6.10.18. Since f is irreducible, we have $t = 1$.

Now set $a_i = \alpha_i^{p^k}$ and write

$$f = \prod_{i=1}^{d} (x^{p^k} - a_i).$$

Then $f(x) = h(x^{p^k})$ for $h = \prod_{i=1}^{d} (x - a_i)$. The polynomial h lies in $F[x]$ since it has the same set of coefficients as f, it is irreducible as any factorization of h would give rise to a factorization of f, and it has α^{p^k} as a root. Also, the a_i are distinct elements, since there are no nontrivial p^kth roots of unity in a field of characteristic p, which tells us that raising to the p^kth power is injective. As E/F is purely inseparable and any root of h generates a separable extension of F, we must have $d = 1$. □

COROLLARY 6.10.21. *Let F be a field of characteristic p, and let E/F be a finite extension. Then $[E : F]_{\mathrm{i}}$ is a power of p.*

We then have the following.

PROPOSITION 6.10.22. *The degree of separability $[K : F]_{\mathrm{s}}$ of a finite extension K/F is equal to the number of embeddings of K fixing F into a given algebraic closure of F.*

PROOF. Let E be the maximal separable subextension of F in K. We know that there are $[K : F]_{\mathrm{s}}$ elements of $\mathrm{Emb}_F(E)$. Any $\alpha \in K - E$ has minimal polynomial $(x-\alpha)^{p^n}$ over E for some $n \geq 1$, so α has only one conjugate over E in K. Thus, any $\varphi \in \mathrm{Emb}_F(E)$ extends uniquely to an embedding of $E(\alpha)$ in $\overline{F}$. Replacing E by $E(\alpha)$ and repeating this last argument, we obtain recursively that φ has a unique extension to all of K. Since every element of $\mathrm{Emb}_F(K)$ is an extension of its restriction to E, the number of such elements is $[K : F]_{\mathrm{s}}$. □

We have the following multiplicativity of separable and inseparable degrees.

LEMMA 6.10.23. *Let K/F be a finite extension and E an intermediate field in K/F. Then*

$$[K : F]_{\mathrm{s}} = [K : E]_{\mathrm{s}}[E : F]_{\mathrm{s}} \quad \text{and} \quad [K : F]_{\mathrm{i}} = [K : E]_{\mathrm{i}}[E : F]_{\mathrm{i}}.$$

PROOF. By the multiplicativity of degrees of field extensions, it suffices to consider separable degrees. It also suffices by recursion to consider this in the case that K can be generated over E by a single element α. Fix an algebraic closure $\bar{F}$ of F. Given a field embedding of E into $\bar{F}$ fixing F, the number of extensions of it to $K = E(\alpha)$ for any $\alpha \in K$ is $[K : E]_{\mathrm{s}}$ by Corollary 6.7.10 and Proposition 6.10.22. The number of such embeddings being $[E : F]_{\mathrm{s}}$, we have the result. □

Finally, we show that finite separable extensions can be generated by a single element.

DEFINITION 6.10.24. We say that a finite field extension E/F is *simple* if there exists $\alpha \in E$ such that $E = F(\alpha)$. In that case, α is said to be a *primitive element* for E/F.

THEOREM 6.10.25 (Primitive element theorem). *Every finite, separable field extension is simple.*

PROOF. Note that if F is finite, then it is isomorphic to $\mathbb{F}_{p^n}$ for some prime p and $n \geq 1$, and by Proposition 6.5.5, it equals $\mathbb{F}_p(\xi)$ for some primitive $(p^n - 1)$th root of unity in F. So we may assume that F is infinite.

Since every finite extension is finitely generated by Corollary 6.2.10, it suffices by recursion to show that if E/F is a finite field extension with $E = F(\alpha, \beta)$ for some $\alpha, \beta \in E$, then there exists $c \in F$ such that $E = F(\alpha + c\beta)$.

Since F is infinite, we can and do choose $c \in F$ such that

$$c \neq -\frac{\alpha' - \alpha}{\beta' - \beta}$$

for all conjugates α' of α over F with $\alpha' \neq \alpha$ and all conjugates β' of β over F with $\beta' \neq \beta$. Set $\gamma = \alpha + c\beta$. Then $\gamma \neq \alpha' + c\beta'$ for all α' and β' as above. Let f be the minimal polynomial of α, and let $h(x) = f(\gamma - cx) \in F(\gamma)[x]$. Then $h(\beta) = f(\alpha) = 0$ and $h(\beta') \neq 0$ for β'. Since h shares the root β with the minimal polynomial g of β over F, but not any other root, and the minimal polynomial q of β over $F(\gamma)$ divides both of the latter polynomials, we must have $q = x - \beta$, which is to say that $\beta \in F(\gamma)$, which then implies that $\alpha \in F(\gamma)$ as well. We therefore have $F(\gamma) = F(\alpha, \beta)$, as desired. □

REMARK 6.10.26. Much as with algebraic closure, we have the notion of a separable closure of a field. A field L is separably closed if it contains a root of every monic, separable polynomial with coefficients in L. Algebraically closed fields are therefore separably closed. A separable closure of a field F is a separable extension F^{sep} of F that is separably closed. If F is a subfield of any separably closed field L, the set of all roots in L of all monic, separable polynomials in $F[x]$ is a subfield that is a separable closure of F. Separable closures exist: in fact, given a field F, take an algebraic closure $\overline{F}$ of F, and it then contains a separable closure F^{sep}, which is the union of all finite separable subextensions of F in $\overline{F}$. Of course, if F is perfect, then the notions of separable closure and algebraic closure of F coincide.

6.11. Normal extensions

We extend the definition of a splitting field to include sets of polynomials.

DEFINITION 6.11.1. Let F be a field, and let S be a subset of $F[x]$ consisting of nonconstant polynomials. A *splitting field* E for S over F is an extension of F such that every polynomial in S splits in E and which contains no proper subextension of F in which this occurs.

EXAMPLE 6.11.2. The field $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is the splitting field of $\{x^2 - 2, x^2 - 3\}$. It is then also the splitting field of $(x^2 - 2)(x^2 - 3)$.

EXAMPLE 6.11.3. An algebraic closure $\overline{F}$ of a field F is a splitting field of the set of all nonconstant polynomials in $F[x]$.

REMARK 6.11.4. An algebraic closure $\overline{F}$ of a field F will always contain a unique splitting field for any subset S of $F[x]$. This field is equal to the intersection of all subfields of $\overline{F}$ in which every polynomial in S splits.

DEFINITION 6.11.5. We say that an algebraic field extension E/F is *normal* if E is the splitting field of some set of polynomials in $F[x]$.

LEMMA 6.11.6. *If E/F is normal, then so is E/F', where F' is any intermediate field in E/F.*

PROOF. If E is the splitting field of a set S of polynomials in $F[x]$, then by definition it contains the splitting field of the set S over F'. If there were a proper subfield of E containing F' in which all the polynomials in S split, then E would not be a splitting field over F, so E is a splitting field of S over F' as well. □

THEOREM 6.11.7. *An algebraic field extension E/F is normal if and only if every field embedding Φ of E that fixes F into an algebraic closure $\overline{F}$ of F containing E satisfies $\Phi(E) = E$. Moreover, under these conditions, E is equal to the splitting field over F of the set of minimal polynomials over F of every element of E.*

PROOF. Suppose first that E/F is normal, and let $S \subseteq F[x]$ be a set of polynomials of which E is a splitting field. Let $\Phi \in \mathrm{Emb}_F(E)$. By definition, E is generated over F by the roots of all polynomials in S. Let $f \in S$, and let $\alpha \in E$ be a root. By Theorem 6.7.9, we must have that $\Phi(\alpha) = \beta$, where β is a root of f in $\overline{F}$. But every root of f in $\overline{F}$ lies in the subfield E, since f splits in E, so $\Phi(\alpha) \in E$. As every element of E may be written as a rational function in the roots of polynomials in S with coefficients in F, we therefore have $\Phi(E) \subseteq E$. Noting Proposition 6.7.15, we then have that $\Phi(E) = E$.

Conversely, suppose that $\Phi(E) = E$ for every $\Phi \in \mathrm{Emb}_F(E)$. Let $\alpha \in E$, and let f be its minimal polynomial over F. Then for any root $\beta \in \overline{F}$ of f, we have an isomorphism $\varphi\colon F(\alpha) \to F(\beta)$ sending α to β. We may then extend the resulting embedding $F(\alpha) \to \overline{F}$ to an embedding $\Phi\colon E \to \overline{F}$. Since $\Phi(E) = E$, we therefore have $\beta \in E$. So E contains the splitting field of every polynomial of F that has a root in E. Since E is algebraic and therefore consists entirely of roots of polynomials in F, it is therefore equal to said splitting field. □

COROLLARY 6.11.8. *Let E/F be a normal field extension, and let $f \in F[x]$ be an irreducible polynomial that has a root in E. Then f splits in E.*

PROOF. This follows directly from the final statement of Theorem 6.11.7. □

For composite extensions, we have the following.

PROPOSITION 6.11.9. *Let F be a field and $\overline{F}$ an algebraic closure of F. Suppose that E and K are subfields of $\overline{F}$ that are normal over F. Then EK/F is normal as well.*

PROOF. We note that any $\varphi \in \mathrm{Emb}_F(EK)$ restricts to embeddings of E and of K into $\overline{F}$. Since E/F and K/F are normal, we have $\varphi(E) = E$ and $\varphi(K) = K$. Every element in EK is a rational function in the elements of $E \cup K$, so $\varphi(EK)$ is contained in EK (and thus equal to EK) as well. By Theorem 6.11.7, EK/F is normal. □

DEFINITION 6.11.10. Let E be a field. An *automorphism* of E is an isomorphism of rings from E to itself.

EXAMPLES 6.11.11.

a. The identity map id_F is an automorphism of any field F, known as the trivial automorphism. It is the identity element in $\mathrm{Aut}(F)$, and it is often denoted by 1.

b. Complex conjugation is an automorphism of $\mathbb{C}$ fixing $\mathbb{R}$.

c. The map $\phi\colon \mathbb{Q}(\sqrt{2}) \to \mathbb{Q}(\sqrt{2})$ sending $a+b\sqrt{2}$ to $a-b\sqrt{2}$ for all $a,b \in \mathbb{Q}$ is an automorphism of $\mathbb{Q}(\sqrt{2})$.

d. The only automorphism of $\mathbb{Q}$ is the trivial automorphism, as the fact that $\phi(1) = 1$ forces $\phi(a) = a$ for all $a \in \mathbb{Q}$ using the properties of a ring homomorphism.

e. The Frobenius map $\varphi_p \colon \overline{\mathbb{F}}_p \to \overline{\mathbb{F}}_p$ defined by $\varphi_p(x) = x^p$ is an automorphism of $\overline{\mathbb{F}}_p$ fixing $\mathbb{F}_p$.

REMARK 6.11.12. The set of automorphisms of a field form a group under composition. That is, the composition of two automorphisms is also an automorphism, as is the inverse of one.

DEFINITION 6.11.13. The *automorphism group* $\mathrm{Aut}(E)$ of a field E is the group of automorphisms of E with the operation of composition.

Often, we are interested in automorphisms fixing a subfield F of E. It is easy to see that these form a subgroup of $\mathrm{Aut}(E)$.

NOTATION 6.11.14. We let $\mathrm{Aut}_F(E)$ denote the subgroup of $\mathrm{Aut}(E)$ for a field E consisting of automorphisms that fix a subfield F.

REMARK 6.11.15. If E is of characteristic 0, then $\mathrm{Aut}_{\mathbb{Q}}(E) = \mathrm{Aut}(E)$.

EXAMPLE 6.11.16. Note that $\mathbb{C} = \mathbb{R}(i)$, and i has minimal polynomial x^2+1. Any automorphism of $\mathbb{C}$ fixing $\mathbb{R}$ must take i to i or $-i$, which then determines the automorphism uniquely. That is, the group $\mathrm{Aut}_{\mathbb{R}}(\mathbb{C})$ consists of exactly two elements, the trivial automorphism and complex conjugation.

The following is an immediate corollary of Theorem 6.11.7, Proposition 6.10.11a, and Proposition 6.10.22.

COROLLARY 6.11.17. *Let E be a finite normal extension of a field F. Then* $\mathrm{Emb}_F(E) = \mathrm{Aut}_F(E)$*, and the order* $[E:F]_{\mathrm{s}}$ *of this group divides* $[E:F]$.

EXAMPLE 6.11.18. Consider the splitting field $E = \mathbb{Q}(\omega, \sqrt[3]{2})$ of x^3-2, where ω is a primitive cube root of unity. Since E is normal, any embedding of E in an algebraic closure of $\mathbb{Q}$ containing E has image E, so gives rise to an automorphism of E. Theorem 6.7.9 then tells us that we may choose such an automorphism uniquely as follows. First, we choose another root of the minimal polynomial x^2+x+1 of ω and send ω to it, i.e., to ω or ω^2. This yields an automorphism of $\mathbb{Q}(\omega)$. Then, we extend this automorphism to an automorphism of E by sending $\sqrt[3]{2}$ to a root of its minimal polynomial over $\mathbb{Q}(\omega)$. Since the degree of $\mathbb{Q}(\omega)$, i.e. 2, is prime to the degree of $\mathbb{Q}(\sqrt[3]{2})$, i.e. 3, over $\mathbb{Q}$, we have $[\mathbb{Q}(\omega, \sqrt[3]{2}) : \mathbb{Q}(\omega)] = 3$, so x^3-2 is still irreducible over $\mathbb{Q}(\omega)$. Therefore, we can send $\sqrt[3]{2}$ to any of $\sqrt[3]{2}$, $\omega\sqrt[3]{2}$, and $\omega^2\sqrt[3]{2}$. That is, there are exactly 6, or $[\mathbb{Q}(\omega, \sqrt[3]{2}) : \mathbb{Q}]$, elements of $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\omega, \sqrt[3]{2}))$.

6.12. Galois extensions

DEFINITION 6.12.1. An algebraic field extension is said to be *Galois* if it is both normal and separable.

REMARK 6.12.2. By Theorem 6.10.19, an algebraic extension of a field of characteristic 0 is Galois if and only if it is normal.

EXAMPLES 6.12.3.

a. The extensions $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(i)$ of $\mathbb{Q}$ are Galois.

b. The extension $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ is not Galois. It is separable but not normal.

c. The extension $\mathbb{F}_p(t^{1/p})/\mathbb{F}_p(t)$ is not Galois. It is normal but not separable.

d. The field $\overline{\mathbb{Q}}$ is a Galois extension of $\mathbb{Q}$.

e. For any $n \geq 1$, the field $\mathbb{F}_{p^n}$ is a Galois extension of $\mathbb{F}_p$.

DEFINITION 6.12.4. Let E/F be a Galois extension. The *Galois group* $\mathrm{Gal}(E/F)$ of E/F is the group of automorphisms of E that fix F.

REMARK 6.12.5. The group $\mathrm{Gal}(E/F)$ is just $\mathrm{Aut}_F(E)$ in our earlier notation. The notation $\mathrm{Gal}(E/F)$ is used only for Galois extensions, whereas $\mathrm{Aut}_F(E)$ can be used for arbitrary extensions.

NOTATION 6.12.6. We often write

$$\begin{tikzcd} E \arrow[d, dash] \\ F \end{tikzcd}$$

to indicate that E is a field extension of F, and if E/F is Galois with Galois group G, we indicate this by the diagram

$$\begin{tikzcd} E \arrow[d, dash, "G"] \\ F. \end{tikzcd}$$

Drawings such as these are known as field diagrams and are useful in illustrating examples.

We will be concerned here only with finite Galois extensions. The following is immediate from Corollary 6.11.17 and Proposition 6.10.11b.

PROPOSITION 6.12.7. *Let E/F be a finite Galois extension of fields. Then* $\mathrm{Gal}(E/F)$ *is a finite group of order* $[E:F]$.

LEMMA 6.12.8. *Let E be a field, and let G be a subgroup of* $\mathrm{Aut}(E)$. *Then the set of elements of E that are fixed by every element of G is a subfield of E.*

PROOF. Let $a, b \in E$ with $b \neq 0$. Let $\sigma \in G$. Then we have

$$\sigma(a-b) = \sigma(a) - \sigma(b) = a - b \quad \text{and} \quad \sigma(ab^{-1}) = \sigma(a)\sigma(b)^{-1} = ab^{-1},$$

so $a-b$ and ab^{-1} are elements of E fixed by G. □

With Lemma 6.12.8 in hand, we may make the following definition.

DEFINITION 6.12.9. Let G be a subgroup of $\mathrm{Aut}(E)$. The *fixed field* E^G of E under G is the largest subfield of E fixed by G.

Note the following.

LEMMA 6.12.10. *Let K/F be a Galois extension, and let E be an intermediate field in K/F. Then K is a Galois extension of E. Moreover, E/F is Galois if and only if it is normal.*

PROOF. The extension K/E is normal by Lemma 6.11.6 and separable by Lemma 6.10.10. The extension E/F is also separable by Lemma 6.10.10, hence the second claim. □

PROPOSITION 6.12.11. *Let K/F be a finite Galois extension. Then the fixed field of K under* $\mathrm{Gal}(K/F)$ *is F.*

PROOF. Let $E = F^{\mathrm{Gal}(K/F)}$. Clearly $F \subseteq E$, and we must show the other containment. By Lemma 6.12.10, the extension K/E is Galois. On the other hand, every element of $\mathrm{Gal}(K/F)$ fixes E, so $\mathrm{Gal}(K/F)$ is equal to its subgroup $\mathrm{Gal}(K/E)$ of automorphisms fixing E. By Proposition 6.12.7, we have that

$$[K:F] = |\mathrm{Gal}(K/F)| = |\mathrm{Gal}(K/E)| = [K:E],$$

which means that $[E:F] = 1$, and therefore $E = F$. □

NOTATION 6.12.12. If K/F is a finite Galois extension and E is an intermediate field, then the restriction of $\sigma \in K$ to an embedding of E into K is denoted $\sigma|_E$.

REMARK 6.12.13. If K/F is a finite Galois extension and E is an intermediate field in K/F such that E/F is Galois, then $\sigma|_E$ is an automorphism of E, so $\sigma|_E \in \mathrm{Gal}(E/F)$.

DEFINITION 6.12.14. Let K/F be a finite Galois extension, and let E be an intermediate field in K/F such that E/F is Galois. Then the *restriction map* from K to E (over F) is the homomorphism of groups $\mathrm{Gal}(K/F) \to \mathrm{Gal}(E/F)$ takes $\sigma \in \mathrm{Gal}(K/F)$ to $\sigma|_E$.

LEMMA 6.12.15. *Let K/F be a Galois extension, and let E be an intermediate field in K/F. Then there exists a bijection of sets*

$$\mathrm{res}_E \colon \mathrm{Gal}(K/F)/\mathrm{Gal}(K/E) \to \mathrm{Emb}_F(E), \qquad \mathrm{res}_E(\sigma\,\mathrm{Gal}(K/E)) = \sigma|_E$$

for $\sigma \in \mathrm{Gal}(K/F)$, where $\overline{F}$ is an algebraic closure of F containing K.

PROOF. Let $\sigma, \tau \in \mathrm{Gal}(K/F)$. We have that $\sigma|_E = \tau|_E$ if and only if $\sigma^{-1}\tau$ fixes E, or equivalently, is an element of $\mathrm{Gal}(K/E)$. In other words, $\sigma|_E = \tau|_E$ if and only if $\sigma\,\mathrm{Gal}(K/E) = \tau\,\mathrm{Gal}(K/E)$. Therefore, res_E is both well-defined and one-to-one.

Given an embedding τ of E into $\overline{F}$ fixing F, we may extend it to an embedding σ of K into $\overline{F}$. Since K/F is normal, σ is an automorphism of K. That is, σ is an element $\mathrm{Gal}(K/F)$ with $\sigma|_E = \tau$, so res_E is surjective. □

PROPOSITION 6.12.16. *Let K/F be a Galois extension, and let E be an intermediate field in K/F. Then E/F is Galois if and only if* $\mathrm{Gal}(K/E)$ *is a normal subgroup of* $\mathrm{Gal}(K/F)$. *If E/F is Galois, then restriction induces an isomorphism*

$$\mathrm{res}_E \colon \mathrm{Gal}(K/F)/\mathrm{Gal}(K/E) \xrightarrow{\sim} \mathrm{Gal}(E/F).$$

PROOF. If E/F is Galois, then the restriction map from $\mathrm{Gal}(K/E)$ to $\mathrm{Gal}(K/F)$ is a surjective homomorphism with kernel exactly $\mathrm{Gal}(K/E)$ by Lemma 6.12.15. So, $\mathrm{Gal}(K/E)$ is normal in $\mathrm{Gal}(K/F)$, and we have the stated isomorphism.

Conversely, suppose that $\mathrm{Gal}(K/E)$ is a normal subgroup of $\mathrm{Gal}(K/F)$. We already know that E/F is separable by Lemma 6.10.10. To show that E/F is normal, it suffices by Theorem 6.11.7 to show that $\varphi(\alpha) \in E$ for all $\alpha \in E$ and field embeddings $\varphi \colon E \to \overline{F}$ fixing F, where

$\overline{F}$ is an algebraic closure of F containing E. Since K/E is Galois, and since E is the fixed field of $\mathrm{Gal}(K/E)$, we have $\varphi(\alpha) \in K$, and we will have $\varphi(\alpha) \in E$ if we can show that $\sigma(\varphi(\alpha)) = \varphi(\alpha)$ for all $\sigma \in \mathrm{Gal}(K/E)$. Since K/F is Galois, we may lift φ to $\tau \in \mathrm{Gal}(K/F)$. The desired equality then amounts to $\sigma\tau(\alpha) = \tau(\alpha)$, or $\tau^{-1}\sigma\tau(\alpha) = \alpha$. Since $\mathrm{Gal}(K/E)$ is normal in $\mathrm{Gal}(K/F)$, we have that $\tau^{-1}\sigma\tau$ fixes E, and in particular α. □

The final ingredient we need is as follows.

PROPOSITION 6.12.17. *Let K/F be a finite Galois extension, and let H be a subgroup of* $\mathrm{Gal}(K/F)$. *Then we have* $\mathrm{Gal}(K/K^H) = H$.

PROOF. By definition, H fixes K^H, so we have $H \leqslant \mathrm{Gal}(K/K^H)$. Since K/F is separable, so is K/K^H, and the primitive element theorem tells us that $K = K^H(\alpha)$ for some $\alpha \in K$. Define

$$f = \prod_{\sigma \in H} (x - \sigma(\alpha)) \in K[x].$$

For $\sigma \in H$, let $\tilde{\sigma}\colon K[x] \to K[x]$ denote the induced homomorphism. We then have $\tilde{\sigma}(f) = f$ for all $\sigma \in H$, which means that $f \in K^H[x]$. In particular, the minimal polynomial of α over K^H divides f, and the degree of that polynomial is $[K : K^H]$, while the degree of f is $|H|$. This implies that $[K : K^H] \leq |H|$, which since $H \leqslant \mathrm{Gal}(K/K^H)$, forces equality on both counts. □

DEFINITION 6.12.18. Let P and Q be sets of subsets of a set X and a set Y, respectively, and suppose that $\phi\colon P \to Q$ is a function. We say that ϕ is *inclusion-reversing* if whenever $A, B \in P$ with $A \subseteq B$, one has $\phi(B) \subseteq \phi(A)$.

We may now state the fundamental theorem of Galois theory, which is essentially just a combination of results we have proven above.

THEOREM 6.12.19 (Fundamental theorem of Galois theory). *Let K/F be a finite Galois extension. Then there are inverse inclusion-reversing bijections*

$$\{\textit{intermediate fields in } K/F\} \underset{\theta}{\overset{\psi}{\rightleftarrows}} \{\textit{subgroups of } \mathrm{Gal}(K/F)\}$$

defined on intermediate fields E in K/F and subgroups H of $\mathrm{Gal}(K/F)$ *by*

$$\psi(E) = \mathrm{Gal}(K/E) \quad \textit{and} \quad \theta(H) = K^H.$$

Moreover, for such E and H, we have

$$[K : E] = |\,\mathrm{Gal}(K/E)| \quad \textit{and} \quad |H| = [K : K^H].$$

These correspondences restrict to bijections

$$\{\textit{normal extensions of } F \textit{ in } K\} \underset{\theta}{\overset{\psi}{\rightleftarrows}} \{\textit{normal subgroups of } \mathrm{Gal}(K/F)\}.$$

Moreover, if E is normal over F (resp., $H \trianglelefteq$ $\mathrm{Gal}(K/F)$*), then restriction induces an isomorphism*

$$\mathrm{Gal}(K/F)/\mathrm{Gal}(K/E) \xrightarrow{\sim} \mathrm{Gal}(E/F) \quad \textit{(resp.,} \ \mathrm{Gal}(K/F)/H \xrightarrow{\sim} \mathrm{Gal}(K^H/F)).$$

PROOF. Let E and H be as in the statement of the theorem. We have that

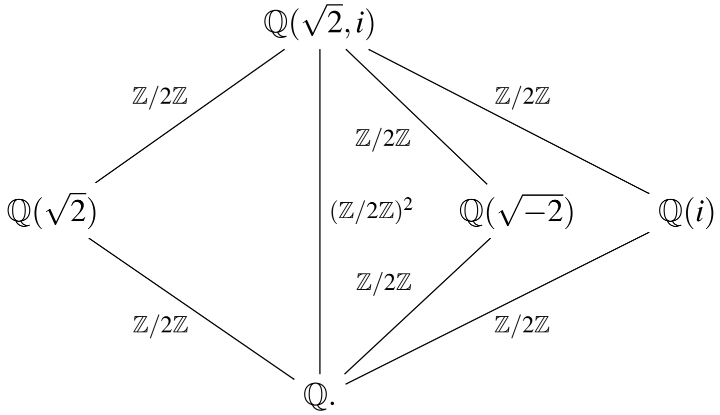
$$\theta(\psi(E)) = \theta(\mathrm{Gal}(K/E)) = K^{\mathrm{Gal}(K/E)} = E$$

by Proposition 6.12.11 and

$$\psi(\theta(H)) = \psi(K^H) = \mathrm{Gal}(K/K^H) = H$$

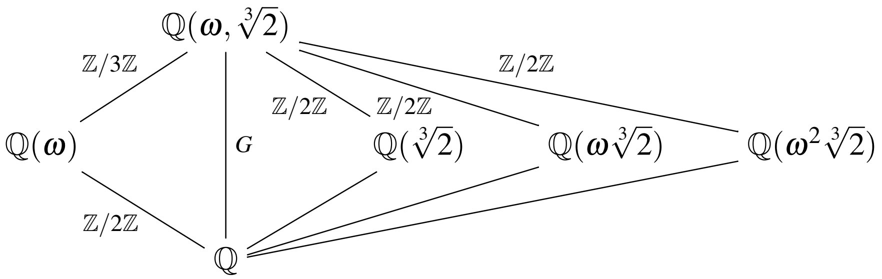
by Proposition 6.12.17, so θ and ψ are inverse bijections. The inclusion-reversing properties of θ and ψ are immediate from the definitions of Galois groups and fixed fields. The statements on orders and indices then follow immediately from Proposition 6.12.7, and the statements on normal extensions and subgroups then become simply Proposition 6.12.16. □

EXAMPLE 6.12.20. The extension $\mathbb{Q}(\sqrt{2},i)/\mathbb{Q}$ is Galois with Galois group isomorphic to the Klein four group. We have the complete field diagram



That is, $\mathrm{Gal}(\mathbb{Q}(\sqrt{2},i)/\mathbb{Q})$ is abelian with two generators σ and τ such that $\sigma(\sqrt{2}) = -\sqrt{2}$, $\sigma(i) = i$, $\tau(\sqrt{2}) = \sqrt{2}$, and $\tau(i) = -i$.

EXAMPLE 6.12.21. Let $G = \mathrm{Gal}(\mathbb{Q}(\omega, \sqrt[3]{2})/\mathbb{Q})$, where ω is a primitive 3rd root of unity. We have the field diagram



As a consequence of the fundamental theorem of Galois theory, we have $G \cong S_3$, since there are only two groups of order 6 up to isomorphism and the cyclic one has a unique subgroup of order 3. It follows that our field diagram contains all of the intermediate fields in $\mathbb{Q}(\omega, \sqrt[3]{2})/\mathbb{Q}$. The fact that the extension $\mathbb{Q}(\sqrt[3]{2})$ is not Galois for $0 \le i \le 2$ corresponds to the fact $\mathrm{Gal}(\mathbb{Q}(\omega, \sqrt[3]{2})/\mathbb{Q}(\sqrt[3]{2}))$

is not a normal subgroup of G, and the other two non-normal intermediate fields correspond to conjugate subgroups.

One can also see this explicitly: note that $\mathrm{Gal}(\mathbb{Q}(\omega,\sqrt[3]{2})/\mathbb{Q}(\omega))$ is generated by an element τ such that $\tau(\sqrt[3]{2}) = \omega\sqrt[3]{2}$, and $\mathrm{Gal}(\mathbb{Q}(\omega,\sqrt[3]{2})/\mathbb{Q}(\sqrt[3]{2}))$ is generated by an element σ such that $\sigma(\omega) = \omega^2$. Then $\tau^3 = 1$, $\sigma^2 = 1$, and

$$\sigma\tau\sigma^{-1}(\omega) = \sigma\tau(\omega^2) = \sigma(\omega^2) = \omega = \tau^{-1}(\omega)$$
$$\sigma\tau\sigma^{-1}(\sqrt[3]{2}) = \sigma\tau(\sqrt[3]{2}) = \sigma(\omega\sqrt[3]{2}) = \omega^2\sqrt[3]{2} = \tau^{-1}(\sqrt[3]{2}),$$

so $\sigma\tau\sigma^{-1} = \tau^{-1}$, and $G = \langle\sigma,\tau\rangle$ is a nonabelian group of order 6, isomorphic to $D_3 \cong S_3$.

More generally, we have the following results on Galois groups of composite fields.

PROPOSITION 6.12.22. *Let L/F be an algebraic extension, and let K and E be extensions of F in L such that K/F is finite Galois. Then EK/E and $K/(E\cap K)$ are finite Galois, and the restriction map*

$$\mathrm{res}_K\colon \mathrm{Gal}(EK/E) \to \mathrm{Gal}(K/(E\cap K)), \qquad \mathrm{res}_K(\sigma) = \sigma|_K \text{ for } \sigma \in \mathrm{Gal}(EK/E)$$

is an isomorphism.

PROOF. First, note that EK/E is normal as it is the splitting field of the same set of polynomials in $F[x]$ that K is over F. Since K/F is finite and separable, we have $K = F(\beta)$ for some $\beta \in K$, so $EK = E(\beta)$, and the fact that the minimal polynomial of β is separable over F tells us that it is over E as well, and therefore EK/E is separable as well. Thus, EK/E is Galois, and $K/(E\cap K)$ is Galois by Lemma 6.12.10.

Now, suppose that $\sigma \in \mathrm{Gal}(EK/E)$ and $\mathrm{res}_K(\sigma) = \sigma|_K = 1$. By definition, we have $\sigma|_E = 1$ as well, so σ fixes every rational function over E in β, and therefore σ fixes EK, which is to say that $\sigma = 1$, or res_K is injective. Now, let H be the image of res_K. The elements of K fixed by H are exactly the elements of K fixed by $\mathrm{Gal}(EK/E)$, so we have

$$K^H = K^{\mathrm{Gal}(EK/E)} = (EK)^{\mathrm{Gal}(EK/E)} \cap K = E\cap K,$$

and therefore $H = \mathrm{Gal}(K/K^H) = \mathrm{Gal}(K/(E\cap K))$, so res_K is surjective as well. □

PROPOSITION 6.12.23. *Let L/F be an algebraic extension, and let K and E be finite Galois extensions of F in L. Then EK/F and $E\cap K/F$ are Galois, and the product of restriction maps*

$$\pi\colon \mathrm{Gal}(EK/F) \to \mathrm{Gal}(K/F)\times\mathrm{Gal}(E/F), \qquad \pi(\sigma) = (\sigma|_K, \sigma|_E) \text{ for } \sigma \in \mathrm{Gal}(EK/F)$$

is an injective homomorphism that is an isomorphism if and only if $E\cap K = F$.

PROOF. That EK/F is separable is Corollary 6.10.12 applied to EK/E and E/F, and that it is normal is Proposition 6.11.9. If $\beta \in E\cap K$, then both E and K contain all roots in $\overline{F}$ of its minimal polynomial, so $E\cap K$ also contains these roots, hence is normal over K. That $E\cap K$ is separable over F follows from the fact that E is.

The kernel of π is exactly those elements of $\mathrm{Gal}(EK/F)$ that fix both K and E, and hence fix all of EK, since every element of EK is a rational function in the elements of E and K. Thus π is injective. Since π is injective, it is surjective if and only if the orders of its domain and codomain are the same, which is to say if and only if

$$[EK:F]=[E:F][K:F].$$

By Proposition 6.12.22, we have

$$[EK:F]=[EK:K][K:F]=[E:E\cap K][K:F],$$

so π is surjective if and only if $E\cap K=F$. □

DEFINITION 6.12.24. Let K/F be a Galois extension.

a. We say that K/F is *abelian* if $\mathrm{Gal}(K/F)$ is abelian.

b. We say that K/F is *cyclic* if $\mathrm{Gal}(K/F)$ is cyclic.

EXAMPLES 6.12.25. We revisit Examples 6.12.20 and 6.12.21.

a. The field $\mathbb{Q}(i,\sqrt{2})$ is the compositum of the normal extensions $\mathbb{Q}(i)$ and $\mathbb{Q}(\sqrt{2})$, which both have Galois group $\mathbb{Z}/2\mathbb{Z}$ and satisfy $\mathbb{Q}(i)\cap\mathbb{Q}(\sqrt{2})=\mathbb{Q}$. By Proposition 6.12.23, we have $\mathrm{Gal}(\mathbb{Q}(i,\sqrt{2})/\mathbb{Q})\cong(\mathbb{Z}/2\mathbb{Z})^2$. The extension $\mathbb{Q}(i,\sqrt{2})/\mathbb{Q}$ is abelian.

b. Take $G=\mathrm{Gal}(\mathbb{Q}(\omega,\sqrt[3]{2})/\mathbb{Q})$. Take $K=\mathbb{Q}(\omega)$ and $E=\mathbb{Q}(\sqrt[3]{2})$. Then $G=\mathrm{Gal}(EK/\mathbb{Q})$, and we set $N=\mathrm{Gal}(EK/K)$ and $H=\mathrm{Gal}(EK/E)$. The map $\mathrm{res}_K\colon G\to\mathrm{Gal}(K/\mathbb{Q})$ is a surjection with kernel N that restricts to an isomorphism on H by Proposition 6.12.22. In particular, H is a complement to N, and G is a semidirect product $N\rtimes H$, nontrivial as $E/\mathbb{Q}$ is not normal. In our case, $N\cong\mathbb{Z}/3\mathbb{Z}$ and $H\cong\mathbb{Z}/2\mathbb{Z}$, so G is nonabelian of order 6, isomorphic to S_3.

The following example is worth being stated as a proposition, as it tells us that all Galois groups of all extensions of finite fields are cyclic.

PROPOSITION 6.12.26. *Let q be a prime power and $n\geq 1$. Then $\mathbb{F}_{q^n}/\mathbb{F}_q$ is cyclic of degree n.*

PROOF. The group $\mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q)$ contains the Frobenius element φ_q with $\varphi_q(\alpha)=\alpha^q$ for all $\alpha\in\mathbb{F}_{q^n}$. For $\varphi_q^r(\alpha)=\alpha^{q^r}$ to equal α would mean that that α is a (q^r-1)th root of unity, which in turn could only happen for all $\alpha\in\mathbb{F}_{q^n}$ if and only if r is a multiple of n. That is, the order of φ_q is n. Therefore, G_n must be cyclic of order n, generated by φ_q. We have that $\mathbb{F}_{q^m}$ is a subfield of $\mathbb{F}_{q^n}$ if and only if m divides n, in which case $\mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_{q^m})=\langle\varphi_q^m\rangle$ is a cyclic group of order n/m. In particular, every finite Galois extension of finite fields is cyclic. □

We can also determine the structure of the Galois groups of cyclotomic extensions of $\mathbb{Q}$. We note that the extension $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ is Galois in that $\mathbb{Q}(\zeta_n)$ is the splitting field of x^n-1.

TERMINOLOGY 6.12.27. For $n\geq 1$ and $a\in(\mathbb{Z}/n\mathbb{Z})^\times$, we will take ζ_n^a to be $\zeta_n^{\tilde{a}}$ for any $\tilde{a}\in\mathbb{Z}$ with $a=\tilde{a}+n\mathbb{Z}$.

DEFINITION 6.12.28. For every $n\geq 1$, the nth *cyclotomic character* is the unique map

$$\chi_n\colon\ \mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})\to(\mathbb{Z}/n\mathbb{Z})^\times$$

such that $\sigma(\zeta_n)=\zeta_n^{\chi_n(\sigma)}$ for all $\sigma\in\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$.

PROPOSITION 6.12.29. *The nth cyclotomic character is an isomorphism for every $n \geq 1$.*

PROOF. We note first that χ_n is a homomorphism. That is, for $\sigma, \tau \in \mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$, we have

$$\zeta_n^{\chi_n(\sigma\tau)} = \sigma\tau(\zeta_n) = \sigma(\zeta_n^{\chi_n(\tau)}) = \sigma(\zeta_n)^{\chi_n(\tau)} = \zeta_n^{\chi_n(\sigma)\chi_n(\tau)}.$$

Next, note that χ_n is injective since an element of $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$ is determined by its value on the generator ζ_n of the extension. Finally, Theorem 6.6.11 implies that $[\mathbb{Q}(\zeta_n) : \mathbb{Q}] = \varphi(n)$, so the orders of the two groups are the same. □

COROLLARY 6.12.30. *The nth cyclotomic field is a finite abelian extension of $\mathbb{Q}$.*

REMARK 6.12.31. The Kronecker-Weber theorem, a proof of which is beyond the scope of these notes, states that every finite abelian extension of $\mathbb{Q}$ is contained inside some cyclotomic field.

6.13. Permutations of roots

We first recall that every finite Galois extension is the splitting field of some polynomial (and in fact we may take that polynomial to be irreducible by the primitive element theorem).

THEOREM 6.13.1. *Let K/F be the splitting field of a separable degree n polynomial in $F[x]$. Then $\mathrm{Gal}(K/F)$ is isomorphic to a subgroup of S_n.*

PROOF. Let K be the splitting field of $f \in F[x]$, and let X be the set of n roots of f. For $\alpha \in X$ and $\sigma \in \mathrm{Gal}(K/F)$, we have $f(\sigma(\alpha)) = \sigma(f(\alpha)) = 0$, so $\sigma(\alpha) \in X$. In other words, $\mathrm{Gal}(K/F)$ acts on X, and thus we have an induced permutation representation $\rho\colon \mathrm{Gal}(K/F) \to S_X$. Note that K is given by adjoining the elements of X to F, so if $\sigma \in \mathrm{Gal}(K/F)$ fixes every element of X, it fixes every element of K and is therefore tirival. Thus, the action of $\mathrm{Gal}(K/F)$ on X is faithful, so ρ is injective. □

COROLLARY 6.13.2. *Let K/F be the splitting field of a separable degree n polynomial in $F[x]$. Then $[K : F]$ divides $n!$.*

EXAMPLES 6.13.3. Again, we revisit Examples 6.12.20 and 6.12.21.

a. The field $\mathbb{Q}(\sqrt{2}, i)$ is the splitting field of $(x^2-2)(x^2+1)$ over $\mathbb{Q}$, which has 4 roots. The image of $\mathrm{Gal}(\mathbb{Q}(\sqrt{2}, i)/\mathbb{Q})$ under any permutation representation on these roots is conjugate to $\langle (1\,2), (3\,4) \rangle$.

b. If we label the roots of x^3-2 in the order $\sqrt[3]{2}$, $\omega\sqrt[3]{2}$, $\omega^2\sqrt[3]{2}$, then we have a permutation representation

$$\rho\colon G = \mathrm{Gal}(\mathbb{Q}(\omega, \sqrt[3]{2})/\mathbb{Q}(\omega)) \xrightarrow{\sim} S_3.$$

We have $G = \langle \sigma, \tau \rangle$ as in Example 6.12.21 with $\rho(\sigma) = (2\,3)$ and $\rho(\tau) = (1\,2\,3)$.

One might ask if every subgroup of S_n, and therefore every finite group, occurs as the Galois group of some extension of fields. As we shall see, the answer is yes.

DEFINITION 6.13.4. Let F be a field, and let $x_1, x_2, \dots, x_n$ be indeterminates. For $1 \le k \le n$, the kth *elementary symmetric polynomial* $s_{n,k}$ in $F[x_1, x_2, \dots, x_n]$ is

$$s_{n,k}(x_1, \dots, x_n) = \sum_{1 \le i_1 < i_2 < \dots < i_k \le n} x_{i_1} x_{i_2} \cdots x_{i_k}.$$

REMARK 6.13.5. Put differently, s_k is the sum over the subsets of $X_n = \{1, 2, \dots, n\}$ of order k of the products of variables with indices in the sets. That is,

$$s_{n,k}(x_1, \dots, x_n) = \sum_{\substack{P \subset X_n \\ |P| = k}} \prod_{i \in P} x_i.$$

As a consequence, $s_{k,n}$ is a sum of $\binom{n}{k}$ monomials.

EXAMPLES 6.13.6. We have $s_{n,1} = x_1 + x_2 + \cdots + x_n$ and $s_{n,n} = x_1 x_2 \cdots x_n$. For $n = 3$, we also have $s_{3,2} = x_1x_2 + x_1x_3 + x_2x_3$, and for $n = 4$, we have

$$s_{4,2} = x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_5 \quad \text{and} \quad s_{4,3} = x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4.$$

PROPOSITION 6.13.7. *The function field $F(x_1, x_2, \dots, x_n)$ is a finite Galois extension of its subfield $F(s_{n,1}, s_{n,2}, \dots, s_{n,n})$, with Galois group isomorphic to S_n.*

PROOF. Let $E = F(s_{n,1}, s_{n,2}, \dots, s_{n,n})$ and $K = \mathbb{Q}(x_1, x_2, \dots, x_n)$. The polynomial

$$f(y) = \prod_{i=1}^{n} (y - x_i) = \sum_{i=0}^{n} s_{n,i} y^i \in F[y]$$

has roots x_i with $1 \le i \le n$. Thus K is the splitting field of f over E. To $\rho \in S_n$, we can associate a unique $\phi(\rho) \in \mathrm{Aut}_F(K)$ by

$$\phi(\rho)(h(x_1, x_2, \dots, x_n)) = h(x_{\rho(1)}, x_{\rho(2)}, \dots, x_{\rho(n)})$$

for $h \in K$. As S_n acts on the set of subsets of X_n of order k, Remark 6.13.5 implies that $\phi(\rho)(s_{n,k}) = s_{n,k}$ for all k, so $\phi(\rho) \in \mathrm{Gal}(K/E)$. The map $\phi \colon S_n \to \mathrm{Gal}(K/E)$ is a homomorphism that is injective by definition and surjective by Theorem 6.13.1. □

We have the following consequence.

COROLLARY 6.13.8. *Every finite group is isomorphic to the Galois group of some field extension.*

PROOF. Let G be a group, and choose n such that H is isomorphic to a subgroup of S_n, which exists by Cayley's theorem. Proposition 6.13.7 yields an extension K/E of fields with $\mathrm{Gal}(K/E) \cong S_n$. Then G is isomorphic to some subgroup H of $\mathrm{Gal}(K/E)$, and we have $H \cong \mathrm{Gal}(K/K^H)$. □

DEFINITION 6.13.9. Let F be a field. The *discriminant* of a monic, degree n polynomial $f \in F[x]$ is

$$\mathrm{D}(f) = \prod_{1 \le i < j \le n} (\alpha_i - \alpha_j)^2,$$

where $f = \prod_{i=1}^{n} (x - \alpha_i)$ in a splitting field of F.

The following lemma is obvious from the definition of the discriminant.

LEMMA 6.13.10. *The discriminant of a monic polynomial f is* 0 *if and only if f is inseparable.*

In fact, the discriminant of a monic polynomial lies in the ground field of the extension, from which it easily follows that it is well-defined independently of the choice of splitting field in its definition.

PROPOSITION 6.13.11. *The discriminant of a monic polynomial $f \in F[x]$ lies in F.*

PROOF. By Lemma 6.13.10, we may suppose that f is separable. Let K be a splitting field of F, and let $\sigma \in \mathrm{Gal}(K/F)$. As σ permutes the roots α_i of f, it induces an element $\rho \in S_n$ such that $\sigma(\alpha_i) = \alpha_{\rho(i)}$. Taking $\Delta = \prod_{1\le i<j\le n}(x_i - x_j)$, we know by Proposition 4.12.1 that $\rho(\Delta) = \pm\Delta$ for the standard action of S_n on polynomials in variables $x_1, x_2, \dots, x_n$. But then $\rho(\Delta^2) = \Delta^2$, so plugging in α_i for x_i, we obtain $\sigma(\mathrm{D}(f)) = \mathrm{D}(f)$. Since $\mathrm{D}(f)$ is fixed by $\mathrm{Gal}(K/F)$, it lies in F. □

REMARK 6.13.12. The proof of Proposition 6.13.11 shows that an element of $\mathrm{Gal}(K/F)$ for the splitting field K of a separable polynomial f of degree n induces an even permutation of the roots of f if and only if it fixes $\prod_{1\le i<j\le n}(\alpha_i - \alpha_j)$.

As a direct consequence of Remark 6.13.12, we have the following.

PROPOSITION 6.13.13. *The discriminant $\mathrm{D}(f)$ of a monic, separable polynomial $f \in F[x]$ is a square in $F^\times$ if and only if the Galois group of its splitting field has image a subgroup of A_n via its permutation representation on the roots of f.*

We explore the consequences of Proposition 6.13.13 for polynomials of low degree.

EXAMPLE 6.13.14. Let $f = x^2 + ax + b \in F[x]$. Let α, β be the roots of F in an algebraic closure of F. The extension $F(\alpha)/F$ is normal, being that it is of degree 1 or 2, so $F(\alpha) = F(\beta)$. Note that $-a = \alpha + \beta$ and $b = \alpha\beta$, so

$$\mathrm{D}(f) = \alpha^2 + \beta^2 - 2\alpha\beta = a^2 - 4b.$$

If $\mathrm{char}\, F = 2$, then $a^2 - 4b = a^2$, so Proposition 6.13.13 tells us that $F(\alpha)/F$ is trivial if $a \neq 0$ and inseparable (possibly trivial) if $a = 0$. If $\mathrm{char}\, F \neq 2$, the extension $F(\alpha)/F$ is separable, so Proposition 6.13.13 again tells us that $a^2 - 4b$ is a square if and only if $\alpha \in F$. This can also be seen by the quadratic formula, which tells us in particular that $F(\alpha) = F(\sqrt{D})$ if $\mathrm{char}\, F \neq 2$.

The case of degree 3 polynomials is rather more involved.

EXAMPLE 6.13.15. Suppose $\mathrm{char}\, F \neq 3$. Let $f = x^3 + ax^2 + bx + c \in F[x]$. Setting $y = x + \frac{a}{3}$, we obtain

$$\begin{aligned} f &= (y - \tfrac{a}{3})^3 + a(y - \tfrac{a}{3})^2 + b(y - \tfrac{a}{3}) + c \\ &= (y^3 - ay^2 + \tfrac{a^2}{3}y - \tfrac{a^3}{27}) + (ay^2 - \tfrac{2a^2}{3}y + \tfrac{a^3}{9}) + (by - \tfrac{ab}{3}) + c \\ &= y^3 + (-\tfrac{a^2}{3} + b)y + (\tfrac{2a^3}{27} - \tfrac{ab}{3} + c). \end{aligned}$$

Set $p = \frac{1}{3}(-a^2+3b)$ and $q = \frac{1}{27}(2a^3 - 9ab + 27c)$, and let $g = x^3 + px + q \in F[x]$.

Let K be a splitting field of f over F, and let $\alpha, \beta, \gamma \in K$ be the roots of g. Then $\alpha+\beta+\gamma = 0$, $s_{3,2}(\alpha,\beta,\gamma) = p$, and $-\alpha\beta\gamma = q$. Note that this implies that

$$0 = (\alpha+\beta+\gamma)^2 = \alpha^2+\beta^2+\gamma^2+2p \tag{6.13.1}$$

and

(6.13.2)

$$p^2 = (\alpha\beta+\alpha\gamma+\beta\gamma)^2 = 2\alpha\beta\gamma(\alpha+\beta+\gamma)+\alpha^2\beta^2+\alpha^2\gamma^2+\beta^2\gamma^2 = \alpha^2\beta^2+\alpha^2\gamma^2+\beta^2\gamma^2.$$

Note that the formal derivative of g is

$$3x^2 + p = s_{3,2}(x-\alpha, x-\beta, x-\gamma),$$

and we can plug α into this, for instance, to obtain

$$3\alpha^2 + p = (\alpha-\beta)(\alpha-\gamma).$$

Doing this also for β and γ and taking the ordering of the differences into account, we obtain by (6.13.1) and (6.13.2) that

$$\begin{aligned}-\mathrm{D}(g) &= (3\alpha^2+p)(3\beta^2+p)(3\gamma^2+p)\\ &= 27\alpha^2\beta^2\gamma^2 + 9p(\alpha^2\beta^2+\alpha^2\gamma^2+\beta^2\gamma^2)+3p^2(\alpha^2+\beta^2+\gamma^2)+p^3\\ &= 27q^2+9p^3-6p^3+p^3 = 27q^2+4p^3.\end{aligned}$$

That is, $\mathrm{D}(g) = -4p^3 - 27q^2$. Since the roots of f and g differ by $\frac{a}{3}$, the differences of the roots of the two are the same, so

$$\mathrm{D}(f) = -4p^3 - 27q^2 = a^2b^2 - 4a^3c + 18abc - 4b^3 - 27c^2.$$

Now, suppose that f is irreducible. Then $\mathrm{Gal}(K/F)$ is isomorphic to a subgroup of S_3 of order divisible by 3, so it is either isomorphic to $A_3 \cong \mathbb{Z}/3\mathbb{Z}$ or S_3, depending on whether $\mathrm{D}(f)$ is a square or not, respectively. If $\mathrm{D}(f) \in F^{\times 2}$, then K is given by adjoining any single root of f. If $\mathrm{D}(f) \notin F^{\times 2}$, then K has a unique intermediate extension $F(\mathrm{D}(f)^{1/2})$ of degree 2, and K is given by adjoining to this any root of f.

We go into a bit less detail for polynomials of degree 4.

EXAMPLE 6.13.16. Let K be the splitting field of a monic, irreducible, separable polynomial f of degree 4 in $F[x]$. If $\mathrm{D}(f) \in F^{\times 2}$, then $\mathrm{Gal}(K/F)$ is isomorphic to a subgroup of A_4 of degree divisible by 4, so A_4 or the Klein 4-group V_4. If $\mathrm{D}(f) \notin F^{\times 2}$, then $\mathrm{Gal}(K/F)$ is isomorphic to $\mathbb{Z}/4\mathbb{Z}$, D_8, or S_4.

Let $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ be the roots of f, and set $\beta_3 = (\alpha_1+\alpha_2)(\alpha_3+\alpha_4)$, $\beta_2 = (\alpha_1+\alpha_3)(\alpha_2+\alpha_4)$, and $\beta_1 = (\alpha_1+\alpha_4)(\alpha_2+\alpha_3)$. The set $\{\beta_1,\beta_2,\beta_3\}$ is a union of orbits under $\mathrm{Gal}(K/F)$, so we can set

$$g(x) = (x-\beta_1)(x-\beta_2)(x-\beta_3) \in F[x]$$

and let E be the splitting field of g over F.

Let $\rho_K \colon \operatorname{Gal}(K/F) \to S_4$ (resp., $\rho_E \colon \operatorname{Gal}(E/F) \to S_3$) be the permutation map for the given ordering of the α_i (resp., β_i). Then we have $\pi \colon S_4 \to S_3$ with kernel $\langle (1\,2)(3\,4), (1\,3)(2\,4) \rangle$ and restricting to the identity on $\langle (1\,2), (1\,2\,3) \rangle$ such that $\pi(\rho_K(\sigma)) = \rho_E(\sigma|_E)$ for all $\sigma \in \operatorname{Gal}(K/F)$.

If g splits, then $\operatorname{Gal}(K/F) \cong V_4$. If g factors as a linear polynomial times an irreducible quadratic, then $\operatorname{Gal}(K/F) \cong D_8$ if f is irreducible over $F(D(f)^{1/2})$ and $\operatorname{Gal}(K/F) \cong \mathbb{Z}/4\mathbb{Z}$ otherwise. If g is irreducible and $\mathrm{D}(g) \in F^{\times 2}$, then $\operatorname{Gal}(E/F) \cong \mathbb{Z}/3\mathbb{Z}$, which forces $\operatorname{Gal}(K/F) \cong A_4$ since 4 divides $[K : F]$. If g is irreducible and $\mathrm{D}(g)$ is not a square in F, then $\operatorname{Gal}(E/F) \cong S_3$, which forces $\operatorname{Gal}(K/F) \cong S_4$.

We next present a proof of the fundamental theorem of algebra that uses Galois theory. We will use the fact that every polynomial of odd degree has a real root (by the intermediate value theorem). We also recall that quadratic polynomials in $\mathbb{C}[x]$ split completely, as is seen via the quadratic formula and the fact that complex numbers have square roots in $\mathbb{C}$.

PROOF OF THE FUNDAMENTAL THEOREM OF ALGEBRA. First, let $f \in \mathbb{C}[x]$ be monic and irreducible, and let $\bar{f} \in \mathbb{C}[x]$ given by applying complex conjugation to its coefficients. The polynomial $g = f\bar{f}$ lies in $\mathbb{R}[x]$ since complex conjugation permutes f and $\bar{f}$, and it suffices to show that g has a root in $\mathbb{C}$. So, we can and do assume that $f \in \mathbb{R}[x]$.

Let $n = \deg f$, and write $n = 2^k m$ for some odd m and $k \geq 0$. If $k = 0$, then f has odd degree and hence a real root, so we suppose $k \geq 1$. By induction, suppose we know that all polynomials in $\mathbb{R}[x]$ of degree 2^{k-1} times an odd number have a root in $\mathbb{C}$. Let $\alpha_1, \alpha_2, \dots, \alpha_n$ be the roots of f in a splitting field Ω of f over $\mathbb{C}$.

For $t \in \mathbb{R}$, define

$$h_t(x) = \prod_{1 \leq i < j \leq n} (x - (\alpha_i + \alpha_j + t\alpha_i\alpha_j)) \in \Omega[x].$$

Any permutation of the α_i's preserves h_t, so $\operatorname{Gal}(\Omega/\mathbb{R})$ fixes h_t, and thus $h_t \in \mathbb{R}[x]$. Note that $\deg h_t = \binom{n}{2} = 2^{k-1}m'$ for some odd m', and thus by induction h_t has a root in $\mathbb{C}$, which necessarily has the form $\alpha_i + \alpha_j + t\alpha_i\alpha_j$ for some $i < j$. In fact, we have such a root for every $t \in \mathbb{R}$, and since that is an infinite set of t, there exist $i < j$ and $s, t \in \mathbb{R}$ such that $\alpha_i + \alpha_j + s\alpha_i\alpha_j$ and $\alpha_i + \alpha_j + t\alpha_i\alpha_j$ are both in $\mathbb{C}$ from which it follows that $\alpha_i + \alpha_j \in \mathbb{C}$ and $\alpha_i\alpha_j \in \mathbb{C}$. But then $(x - \alpha_i)(x - \alpha_j) \in \mathbb{C}[x]$, which being quadratic, has a root in $\mathbb{C}$. □

Part 2

A Second Course

CHAPTER 7

Topics in group theory

7.1. Semidirect products

PROPOSITION 7.1.1. *Let N and H be groups and $\varphi\colon H \to \mathrm{Aut}(N)$ be a homomorphism. Then there exists a group G with underlying set $N \times H$ and group operation*

$$(n,h)\cdot(n',h') = (n\varphi(h)(n'),hh')$$

for all $n,n' \in N$ and $h,h' \in H$. Moreover, $H = \{e\} \times H \leqslant G$ and $N = N \times \{e\} \trianglelefteqslant G$. In fact, in G we have $\varphi(h)(n) = hnh^{-1}$ for all $h \in H$ and $n \in N$.

PROOF. We note that $(e,e) \in G$ is an identity, that $(\varphi(h^{-1})(n^{-1}),h^{-1})$ is inverse to (n,h), and we leave it to the reader to check associativity. Clearly $H,N \leqslant G$ by definition of the multiplication, and we check that for $h \in H$ and $n \in N$, we have

$$hnh^{-1} = (e,h)(n,e)(e,h^{-1}) = (e,h)(n,h^{-1}) = (\varphi(h)(n),e) = \varphi(h)(n) \in N.$$

□

DEFINITION 7.1.2. For groups N and H and a homomorphism $\varphi\colon H \to \mathrm{Aut}(N)$, the group defined by Proposition 7.1.1 is known as the *semidirect product* of N and H relative to φ and is denoted by $N \rtimes_\varphi H$.

EXAMPLE 7.1.3. If H and N are groups and $\varphi\colon H \to \mathrm{Aut}(N)$ satisfies $\varphi(h) = \mathrm{id}_N$ for all $h \in H$, then $H \rtimes_\varphi N$ is the direct product $H \times N$.

EXAMPLE 7.1.4. Let $\varphi\colon (\mathbb{Z}/n\mathbb{Z})^\times \to \mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$ be the isomorphism taking $a \in (\mathbb{Z}/n\mathbb{Z})^\times$ to multiplication by a. Set $G = \mathbb{Z}/n\mathbb{Z} \rtimes_\varphi (\mathbb{Z}/n\mathbb{Z})^\times$. Then $G \xrightarrow{\sim} \mathrm{Aff}(\mathbb{Z}/n\mathbb{Z})$ via $(b,a) \mapsto \left(\begin{smallmatrix} a & b \\ 0 & 1 \end{smallmatrix}\right)$, so G is also isomorphic to $\mathrm{Aut}(D_n)$ by Proposition 4.3.5.

PROPOSITION 7.1.5. *Let G be a group with normal subgroup N and subgroup H such that $N \cap H = \{e\}$ and $NH = G$. Define a homomorphism $\varphi\colon H \to \mathrm{Aut}(N)$ by $\varphi(h)(n) = hnh^{-1}$. Then we may define an isomorphism of groups by*

$$\psi\colon N \rtimes_\varphi H \to G, \qquad \psi(n,h) = nh$$

for all $n \in N$ and $h \in H$.

PROOF. Any $g \in G$ can be written as nh for some $n \in N$ and $h \in H$ by assumption, so f is onto. For $n,n' \in N$ and $h,h' \in H$, we have

$$\psi((n,h)(n',h')) = \psi(n\varphi(h)(n'),hh') = n\varphi(h)(n')hh' = nhn'h' = \psi((n,h))f((n',h')).$$

If $\psi(n,h) = nh = e$, then $n = h^{-1} \in N \cap H$, so $n = h = e$. □

The proposition we have just proven has Proposition 4.11.4 as a corollary.

ALTERNATE PROOF OF PROPOSITION 4.11.4. By Proposition 7.1.5, we need only that $\varphi\colon H \to \mathrm{Aut}(N)$ given by $\varphi(h)(n) = hnh^{-1}$ for $h \in H$ and $n \in N$ is the trivial map. That is, we need that $hnh^{-1} = n$, or $[h,n] = e$, for all such h and n. This follows as H and N are normal, so $[h,n] \in H \cap N = \{e\}$. □

DEFINITION 7.1.6. If G is a group with subgroups N and H such that $G \cong N \rtimes_\varphi H$ for $\varphi\colon H \to \mathrm{Aut}(N)$ given by $\varphi(h)(n) = hnh^{-1}$, then we say that G is the *internal semidirect product* of N and H and write $G = N \rtimes H$ to denote this.

DEFINITION 7.1.7. Let G be a group with normal subgroup N. A *complement* to N in G is a subgroup H such that G is the internal semidirect product $N \rtimes H$ of N and H.

REMARK 7.1.8. There are often many complements to a normal subgroup. In particular, if $G = N \rtimes H$ and $n \in N$, then nHn^{-1} is also a complement to N. If N is abelian, then we have the equality $\gamma_{nhn^{-1}} = \gamma_h$ of conjugation maps, but if N is nonabelian, then these may not be equal. The map $\varphi\colon H \to \mathrm{Aut}(N)$ given by $\varphi(h) = \gamma_{nhn^{-1}}$ would then satisfy $G \cong N \times_\varphi H$, though not necessarily internally. Rather, this is simply an expression of the fact that $G = N \rtimes nHn^{-1}$.

The following rather general result can be used to show that two semidirect products are isomorphic.

PROPOSITION 7.1.9. *Let H, H', N, and N' be groups and $\varphi\colon H \to \mathrm{Aut}(N)$ and $\varphi'\colon H' \to \mathrm{Aut}(N')$ be homomorphisms. Suppose that there exist isomorphisms $\psi\colon H \to H'$ and $\theta\colon N \to N'$, and define*

$$\Theta\colon \mathrm{Aut}(N) \to \mathrm{Aut}(N')$$

by $\Theta(\alpha) = \theta \circ \alpha \circ \theta^{-1}$ for any $\alpha \in \mathrm{Aut}(N)$. If $\Theta \circ \varphi = \varphi' \circ \psi$, then the map

$$f\colon N \rtimes_\varphi H \to N' \rtimes_{\varphi'} H'$$

defined by $f(n,h) = (\theta(n), \psi(h))$ for all $n \in N$ and $h \in H$ is an isomorphism.

PROOF. Note that f has an inverse given by $f^{-1}(n',h') = (\theta^{-1}(n'), \psi^{-1}(h'))$ for all $n' \in N$ and $h' \in H$, so we need only show that f is a homomorphism. Letting $n_1, n_2 \in N$ and $h_1, h_2 \in H$, we calculate:

$$f((n_1,h_1)(n_2,h_2)) = f(n_1\varphi(h_1)(n_2), h_1h_2) = (\theta(n_1)\theta(\varphi(h_1)(n_2)), \psi(h_1)\psi(h_2)),$$
$$f(n_1,h_1)f(n_2,h_2) = (\theta(n_1), \psi(h_1))(\theta(n_2), \psi(h_2)) = (\theta(n_1)\varphi'(\psi(h_1))(\theta(n_2)), \psi(h_1)\psi(h_2)).$$

To see that the first coordinates of these expressions are equal, we check that

$$\varphi'(\psi(h_1))(\theta(n_2)) = \Theta(\phi(h_1))(\theta(n_2)) = (\theta \circ \phi(h_1) \circ \theta^{-1})(\theta(n_2)) = \theta(\phi(h_1)(n_2)).$$

Thus, f is a homomorphism. □

We can use this to completely classify groups of order a product of two distinct primes, completing the study begun in Theorem 4.11.5.

THEOREM 7.1.10. *Let p and q be distinct primes with $q \equiv 1 \bmod p$. Then there exists a unique isomorphism class of nonabelian groups of order pq.*

PROOF. Let G be a nonabelian group. By Theorem 4.11.5, we have that it has a unique normal subgroup Q of order q, and let P be a subgroup of order p. By Proposition 7.1.5, we have that $G = Q \rtimes P$. We have $P \cong \mathbb{Z}/p\mathbb{Z}$ and $Q \cong \mathbb{Z}/q\mathbb{Z}$. Fixing such isomorphisms and recalling the canonical isomorphism $\mathrm{Aut}(\mathbb{Z}/q\mathbb{Z}) \cong (\mathbb{Z}/q\mathbb{Z})^\times$, we are reduced to showing that there is a unique isomorphism class of semi-direct product $(\mathbb{Z}/q\mathbb{Z}) \rtimes_\varphi (\mathbb{Z}/p\mathbb{Z})$, where $\varphi \colon \mathbb{Z}/p\mathbb{Z} \to (\mathbb{Z}/q\mathbb{Z})^\times$ is a nontrivial homomorphism. The group $(\mathbb{Z}/q\mathbb{Z})^\times$ is cyclic by Corollary 6.5.5. Let $a \in (\mathbb{Z}/q\mathbb{Z})^\times$ be a generator.

Any nontrivial homomorphism $\varphi \colon \mathbb{Z}/p\mathbb{Z} \to (\mathbb{Z}/q\mathbb{Z})^\times$ must send 1 to an element of order p in $(\mathbb{Z}/q\mathbb{Z})^\times$. If we set $b = a^{(q-1)/p}$, then $\varphi(1) = b^i$ for some $i \in \mathbb{Z}$ with $i \not\equiv 0 \bmod p$. Let us denote this particular homomorphism by φ_i, and define $\psi_i \colon \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ to be multiplication by i. Then $\varphi_i = \varphi_1 \circ \psi_i$ since both maps send 1 to b^i. Proposition 7.1.9 then tells us that the semidirect products $(\mathbb{Z}/q\mathbb{Z}) \rtimes (\mathbb{Z}/p\mathbb{Z})$ defined by φ_1 and φ_i are isomorphic. That is, there is a unique isomorphism class of nonabelian semidirect product of order pq. □

7.2. Composition series

First, we explain how simple groups may be used in building arbitrary finite groups, starting with the following definition.

DEFINITION 7.2.1. Any collection $(H_i)_{i\in\mathbb{Z}}$ of subgroups of a group G with $H_{i-1} \leqslant H_i$ for $i \in \mathbb{Z}$ is called a *series of subgroups of G*.

DEFINITION 7.2.2. Let $\mathscr{C} = (H_i)_{i\in\mathbb{Z}}$ be a series of subgroups of a group G.

a. We say that $\mathscr{C}$ is an *ascending series* if $H_i = 1$ for i sufficiently small.

b. We say that $\mathscr{C}$ is a *descending series* if $H_i = G$ for i sufficiently large.

c. We say that $\mathscr{C}$ is a *finite series* if it is both ascending and descending.

d. The *length* of a finite series $\mathscr{C}$ is difference $j - i$ of the smallest integer j such that $H_j = G$ and largest integer i such that $H_i = 1$.

NOTATION 7.2.3. We use the notation

$$1 = H_0 \leqslant H_1 \leqslant \cdots \leqslant H_{t-1} \leqslant H_t = G$$

to denote a finite series of subgroups H_i of a group G with $H_0 = 1$, $H_t = G$. It has length t if $H_0 \neq H_1$ and $H_{t-1} \neq H_t$.

REMARK 7.2.4. To say that a series $(H_i)_{i\in\mathbb{Z}}$ of subgroups of G is finite is stronger than simply saying it has only finitely many terms. For instance, if G is nontrivial, then $H_i = 1$ for all $i \in \mathbb{Z}$ provides a series with only one distinct subgroup, but it is not finite as no H_i equals G.

REMARK 7.2.5. A descending series in G is often taken to be a list $(H_i)_{i\in\mathbb{Z}}$ of subgroups of G with $H_i \leqslant H_{i-1}$ for all i and $H_i = G$ for i sufficiently small. This agrees with the usual notion in the sense that letting $K_i = H_{-i}$ will provide a descending series $(K_i)_{i\in\mathbb{Z}}$ in the sense of the original definition.

DEFINITION 7.2.6. A finite series

$$1 = H_0 \leqslant H_1 \leqslant \cdots \leqslant H_{t-1} \leqslant H_t = G$$

of subgroups of G is said to be a subnormal series if $H_{i-1} \trianglelefteq H_i$ for all $1 \le i \le t$. It is called a normal series if $H_i \trianglelefteq G$ for all $0 \le i \le t-1$.

DEFINITION 7.2.7. Two subnormal series $(H_i)_{i=0}^t$ and $(K_i)_{i=0}^t$ are equivalent if there exists $\sigma \in S_t$ such that $H_i/H_{i-1} \cong K_{\sigma(i)}/K_{\sigma(i)-1}$ for all $1 \le i \le t$.

DEFINITION 7.2.8. A *refinement* of a subnormal series $(H_i)_{i=0}^t$ in a group G is a subnormal series $(K_i)_{i=0}^s$ such that there exists an increasing function $f\colon \{0,\dots,t\} \to \{0,\dots,s\}$ such that $H_i = K_{f(i)}$ for $0 \le i \le t$.

THEOREM 7.2.9 (Schreier refinement theorem). *Any two subnormal series in a group G have refinements that are equivalent.*

PROOF. Let $(H_i)_{i=0}^t$ and $(K_i)_{i=0}^s$ be subnormal series in G. For $0 \le i < t$ and $0 \le j < s$, let

$$M_{si+j} = H_i(H_{i+1} \cap K_j) \quad \text{and} \quad N_{tj+i} = K_j(K_{j+1} \cap H_i).$$

Set $M_{mn} = N_{mn} = G$ as well. Then $M_{si+j} \trianglelefteq M_{si+j+1}$ for $0 \le i \le t-1$ and $0 \le j \le s-2$ as $K_j \trianglelefteq K_{j+1}$, and $M_{si+s-1} \trianglelefteq H_{i+1} = M_{s(i+1)}$ for $0 \le i \le t-1$ as $H_i \trianglelefteq H_{i+1}$ and $H_{s-1} \trianglelefteq G$. Thus $(M_i)_{i=0}^{mn}$ is a subnormal series, as is $(N_i)_{i=0}^{mn}$. In fact, we see from this that $(M_i)_i$ refines $(H_i)_i$ and $(N_i)_i$ refines $(K_i)_i$.

It remains to see that $(M_i)_i$ and $(N_i)_i$ are equivalent. For $0 \le i \le t-1$ and $0 \le j \le s-2$, note that

$$\frac{M_{si+j+1}}{M_{si+j}} \cong \frac{H_i(H_{i+1} \cap K_{j+1})}{H_i(H_{i+1} \cap K_j)} \cong \frac{K_j(K_{j+1} \cap H_{i+1})}{K_j(K_{j+1} \cap H_i)} \cong \frac{N_{tj+i+1}}{N_{tj+i}}$$

by the butterfly lemma, and

$$\frac{M_{s(i+1)}}{M_{s(i+1)-1}} \cong \frac{H_{i+1}}{H_i(H_{i+1} \cap K_{s-1})} \cong \frac{K_{s-1}H_{i+1}}{K_{s-1}H_i} \cong \frac{K_{s-1}(K_s \cap H_{i+1})}{K_{s-1}(K_s \cap H_i)} \cong \frac{K_{(s-1)t+i+1}}{K_{(s-1)t+i}}$$

for $0 \le i \le t-1$ since $K_s = G$. Thus, the two refinements are equivalent. □

DEFINITION 7.2.10. A subnormal series of subgroups

$$1 = H_0 \lhd H_1 \lhd \cdots \lhd H_{t-1} \lhd H_t = G$$

of a group G is called a *composition series* for G if H_i/H_{i-1} is simple for each $1 \le i \le t$. The simple groups H_i/H_{i-1} are referred to as the *composition factors* of the series.

LEMMA 7.2.11. *Let*

$$1 = H_0 \lhd H_1 \lhd \cdots \lhd H_{t-1} \lhd H_t = G$$

be a composition series for G, and let N be a proper normal subgroup of G.

a. There exists $s \le t$ and an increasing function $f\colon \{0,\dots,s\} \to \{0,\dots,t\}$ with $f(0) = 0$ such that

$$1 \lhd H_{f(1)} \cap N \lhd H_{f(2)} \cap N \lhd \cdots \lhd H_{f(s)} \cap N = N$$

is a composition series for N with composition factors

$$\frac{H_{f(i)} \cap N}{H_{f(i-1)} \cap N} \cong \frac{H_{f(i)}}{H_{f(i)-1}}.$$

b. Set $\bar{H}_i = H_i/(H_i \cap N)$ for $0 \le i \le t$. There exists an $r \le t$ and an increasing function $f'\colon \{0,1,\ldots,r\} \to \{0,1,\ldots,t\}$ with $f(0) = 0$ such that

$$1 = \bar{H}_0 \lhd \bar{H}_{f'(1)} \lhd \cdots \lhd \bar{H}_{f'(r-1)} \lhd \bar{H}_{f'(r)} = G/N$$

is a composition series for G/N with composition factors

$$\bar{H}_{f'(i)}/\bar{H}_{f'(i-1)} \cong H_{f'(i)}/H_{f'(i)-1}.$$

c. In the notation of parts a and b, the images of f and f' to be complementary away from 0,x and $r+s$ to equal t.

Proof. Let $0 \le i \le t-1$. The quotient $(H_i \cap N)/(H_{i-1} \cap N)$ is a subgroup of the simple group H_i/H_{i-1} and therefore necessarily trivial or improper. Let s be the number of simple quotients. Let $f(0) = 0$, and for $1 \le j \le s$, let $f(j)$ be the smallest positive integer greater than $f(j-1)$ and such that $(H_{f(j)} \cap N)/(H_{f(j)-1} \cap N)$ is simple. Then $H_{f(j)-1} \cap N = H_{f(j-1)} \cap N$, and the result follows.

Similarly, by the third isomorphism theorem, we have

$$\frac{\bar{H}_i}{\bar{H}_{i-1}} \cong \frac{H_i}{H_{i-1}(H_i \cap N)},$$

which is a quotient of H_i/H_{i-1} by the image of $H_{i+1} \cap N$ in it. Since H_{i+1}/H_i is simple, this image is either trivial or H_{i+1}/H_i. That is, $\bar{H}_{i+1}/\bar{H}_i$ is either trivial or simple. Let s be the number of simple terms. Set $f'(0) = 0$, and for $1 \le j \le r$, take $f'(j)$ to be the smallest integer greater than $f'(j-1)$ such that $\bar{H}_{f'(j)}/\bar{H}_{f'(j)-1}$ is simple. Then $\bar{H}_{f'(j)-1} = \bar{H}_{f'(j-1)}$.

Note that $\bar{H}_i \ne \bar{H}_{i-1}$ if and only if $H_i \cap N = H_{i-1} \cap N$. Then $r+s = t$ and the images of f and f' are complementary by construction. □

We leave the straightforward proof of the following lemma to the reader.

Lemma 7.2.12. *Let G be a group, and let N be a normal subgroup. Suppose that N has a composition series*

$$1 = H_0 \lhd H_1 \lhd \cdots \lhd H_{t-1} \lhd H_s = N$$

and G/N has a composition series

$$1 = Q_0 \lhd Q_1 \lhd \cdots \lhd Q_{r-1} \lhd Q_r = G/N.$$

For $1 \le i \le r$, let H_{s+i} denote the unique subgroup of G containing N and such that $H_{s+i}/N = Q_i$, which exists by Proposition 2.13.10. Then the series

$$1 = H_0 \lhd H_1 \lhd \cdots \lhd H_{t-1} \lhd H_t = G$$

is a composition series of G with composition factors satisfying $H_{s+i}/H_{s+i-1} = Q_i/Q_{i-1}$ for $1 \le i \le r$.

COROLLARY 7.2.13. *Let G be a group and N a normal subgroup. If N and G/N have composition series, then G has a composition series. Moreover, its list of composition factors consists of the concatenation of the list of composition factors of N by the list of composition factors of G.*

THEOREM 7.2.14 (Jordan-Hölder theorem).

a. Every finite group has a composition series.

b. Let G be a nontrivial group with composition series

$$1 = N_0 \lhd N_1 \lhd \cdots \lhd N_{s-1} \lhd N_s = G$$

and

$$1 = H_0 \lhd H_1 \lhd \cdots \lhd H_{t-1} \lhd H_t = G.$$

Then $s = t$ and there exists a permutation $\sigma \in S_t$ such that

$$H_{\sigma(i)}/H_{\sigma(i)-1} \cong N_i/N_{i-1}$$

for all $1 \le i \le t$.

PROOF. To show part a, we work by induction on the order n of the group G. It is clear in the case that G is trivial, with $t = 0$. Now, if G is nontrivial of order n, then either it is simple, and the composition series is $1 \leqslant G$, or it is not, and there exists a nontrivial normal subgroup $K \lhd G$, and then K and G/K have composition series by induction. The result is then immediate from Lemma 7.2.13.

To see that the composition series is unique in the stated sense of part b, start with two composition series as in the statement of the theorem. We work by induction on the minimal length s of a composition series for G. If $s = 0$, then G is trivial. If $s = 1$, then G is simple, so it cannot have a nontrivial normal subgroup, and all composition series must have length 1. Consider $N = N_{s-1}$, which has the composition series

$$1 \lhd N_1 \lhd \cdots \lhd N_{s-2} \lhd N_{s-1} = N,$$

as well as a composition series

$$1 \lhd H_{f(1)} \cap N \lhd \cdots \lhd H_{f(r-1)} \cap N \lhd H_{f(r)} \cap N = N$$

for some $r \le t$ and increasing $f\colon X_r \to X_t$ by Lemma 7.2.11a. Since the minimal length of a composition series of N is less than s, we have by induction that $r = s-1$ and there exists $\sigma \in S_{s-1}$ such that

$$N_i/N_{i-1} \cong (H_{f(\sigma(i))} \cap N)/(H_{f(\sigma(i)-1)} \cap N) \cong H_{f(\sigma(i))}/H_{f(\sigma(i))-1}$$

for all i, again by Lemma 7.2.11a.

Let $k < t$ be maximal such that $H_{k-1} \leqslant N$. Then

$$H_{k-1} \cap N = H_{k-1} \leqslant H_k \cap N < H_k.$$

Since H_k/H_{k-1} is simple, this forces $H_{k-1} = H_k \cap N$. In particular, k is not in the image of f. Moreover, we have

$$H_k/H_{k-1} \cong H_k/(H_k \cap N) \cong H_kN/N \cong G/N,$$

the latter step as N is a maximal normal subgroup of G and $H_k \not\leqslant N$. As we have found the final composition factor in the series $(N_i)_i$ among those of the series $(H_i)_i$, it remains only to show that $s = t$. If $(H_i \cap N)/(H_{i-1} \cap N)$ nontrivial for any for $i \neq k$, then by Lemma 7.2.11c, the group G/N has a composition series of length at least 2, but G/N is simple, so this is impossible. Thus, $r = t - 1$ as well, as needed. □

DEFINITION 7.2.15. The *Jordan-Hölder factors* of a group G are the terms in a list of the isomorphism classes of the composition factors in a composition series for G.

EXAMPLES 7.2.16.

a. The group $\mathbb{Z}/p^n\mathbb{Z}$ for a prime p has n copies of $\mathbb{Z}/p\mathbb{Z}$ as its Jordan-Hölder factors, which arise from its unique composition series

$$0 \lhd \langle p^{n-1} \rangle \lhd \langle p^{n-2} \rangle \lhd \cdots \lhd \langle p \rangle \lhd \mathbb{Z}/p^n\mathbb{Z}.$$

b. The group $\mathbb{Z}/6\mathbb{Z}$ has two composition series

$$0 \lhd \langle 2 \rangle \lhd \mathbb{Z}/6\mathbb{Z} \quad \text{and} \quad 0 \lhd \langle 3 \rangle \lhd \mathbb{Z}/6\mathbb{Z},$$

both of which have Jordan-Hölder factors $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$.

c. Let G be a nonabelian group order pq with p and q distinct primes and $q \equiv 1 \bmod p$. Then G has a unique composition series $1 \lhd Q \lhd G$, where Q has order q, and it has Jordan-Hölder factors $\mathbb{Z}/p\mathbb{Z}$ and $\mathbb{Z}/q\mathbb{Z}$.

d. For $n \geq 6$, the group S_n has a unique composition series $1 \lhd A_n \lhd G$ with Jordan-Hölder factors A_n and $\mathbb{Z}/2\mathbb{Z}$.

REMARK 7.2.17. The set of Jordan-Hölder factors of a group tell us a great deal about the structure of a group, but they do not tell us the group. For instance, $\mathbb{Z}/n^2\mathbb{Z}$ and $(\mathbb{Z}/n\mathbb{Z})^2$ have the same Jordan-Hölder factors for any $n \geq 2$.

7.3. Solvable groups

DEFINITION 7.3.1. Let G be a group. The *derived series* of G is the unique descending series $(G^{(i)})_{i \geq 0}$ of subgroups of G with $G^{(0)} = G$ and $G^{(i)} = [G^{(i-1)}, G^{(i-1)}]$ for all $i \geq 1$.

NOTATION 7.3.2. Often, one writes G' for $G^{(1)} = [G, G]$ and G'' for $G^{(2)} = [[G, G], [G, G]]$.

DEFINITION 7.3.3. A group G is *solvable* if its derived series is finite.

EXAMPLES 7.3.4.

a. The derived series of an abelian group satisfies $G^{(i)} = 1$ for all $i \geq 1$. Hence, abelian groups are solvable.

b. The derived series of a nonabelian simple group G satisfies $G^{(i)} = G$ for $i \geq 0$. Hence, nonabelian simple groups are not solvable.

EXAMPLE 7.3.5. Let R be a commutative ring. Consider the group

$$T = \mathrm{Heis}(R) = \left\{ \left(\begin{smallmatrix} 1 & a & c \\ & 1 & b \\ & & 1 \end{smallmatrix} \right) \mid a, b, c \in R \right\} \leqslant \mathrm{GL}_3(R).$$

The reader should verify that this group has commutator subgroup equal to its center, which is

$$Z(T) = [T,T] = \left\{ \begin{pmatrix} 1 & 0 & c \\ & 1 & 0 \\ & & 1 \end{pmatrix} \mid c \in R \right\} \leqslant \mathrm{GL}_3(R).$$

In particular, $T^{(2)} = [Z(T), Z(T)] = 0$. In fact, the reader might find a rather canonical isomorphism from $\mathrm{Heis}(\mathbb{Z})$ to the group presented by $\langle x, y, z \mid [x,y] = z, [x,z] = [y,z] = e \rangle$.

LEMMA 7.3.6. *The groups $G^{(i)}$ for $i \geq 1$ are characteristic subgroups of a group G.*

PROOF. First, Lemma 4.3.15c tells us that $G^{(i)}$ is characteristic in $G^{(i-1)}$ for each $i \geq 1$, and then the result follows recursively from Lemma 4.3.16. □

We can now prove the following equivalence of definitions of solvability.

PROPOSITION 7.3.7. *The following statements regarding a group G are equivalent:*

i. G is solvable,

ii. G has a normal series with abelian composition factors, and

iii. G has a subnormal series with abelian composition factors.

PROOF. That (i) implies (ii) is a consequence of the facts that the group $G^{(i)}$ are characteristic, hence normal, and that $G^{(i-1)}/G^{(i)}$ is the quotient of $G^{(i-1)}$ by its commutator subgroup, hence abelian. That (ii) implies (iii) is obvious. So, suppose (iii) and let

$$G = N_0 \rhd N_1 \rhd \cdots \rhd N_{t-1} \rhd N_t = 1$$

be a subnormal series of length t. (Note the reversed indexing, as in Remark 7.2.5.) We claim that $G^{(i)} \leqslant N_i$ for each $i \geq 0$. For $i = 0$, we have $G = G^{(0)} = N_0$. In general, suppose inductively that $G^{(i-1)} \leqslant N_{i-1}$. Then $G^{(i)} \leqslant [N_{i-1}, N_{i-1}]$ by definition, and we have $[N_{i-1}, N_{i-1}] \leqslant N_i$ as N_{i-1}/N_i is abelian. Therefore, we have that $G_t = N_t = 1$, and G is solvable. □

We also have the following.

PROPOSITION 7.3.8.

a. Every subgroup of a solvable group is solvable.

b. Every quotient group of a solvable group is solvable.

c. If G is a group and N is a normal subgroup of G such that N and G/N are both solvable, then G is solvable as well.

PROOF. Let G be a group and N a normal subgroup. If G is solvable, then it has a composition series with abelian factors, so N and G/N are solvable by Lemma 7.2.11. Part (iii) is a corollary of Corollary 7.2.13, since the derived series of N and G/N have abelian composition factors. □

PROPOSITION 7.3.9. *A group G with a composition series is solvable if and only if it is finite and its Jordan-Hölder factors are all cyclic of prime order.*

PROOF. If G has cyclic Jordan-Hölder factors, then G is solvable by Proposition 7.3.7. If G is solvable and has a composition series, then the composition factors are abelian by Proposition 7.3.7 and the uniqueness in Theorem 7.2.14. As composition factors, they are also simple, hence cyclic of prime order, from which it follows that G is finite. □

EXAMPLE 7.3.10. All groups of order pq for distinct primes p and q are solvable, as their Jordan-Hölder factors are $\mathbb{Z}/p\mathbb{Z}$ and $\mathbb{Z}/q\mathbb{Z}$.

DEFINITION 7.3.11. A *Hall subgroup* of a finite group G is a subgroup H such that $|H|$ and $[G:H]$ are relatively prime.

7.4. Nilpotent groups

DEFINITION 7.4.1. Let G be a group. The *lower central series* of G is the unique descending series $(G_i)_{i\geq 1}$ of G with $G_1 = G$ and

$$G_{i+1} = [G, G_i] = \langle \{[a,b] \mid a \in G, b \in G_i\} \rangle$$

for each $i \geq 1$.

REMARK 7.4.2. By convention, G_i starts with $G_1 = G$, while $G^{(i)}$ starts with $G^{(0)} = G$.

REMARK 7.4.3. For a group G, we have $G' = G_2$, but $G'' = [G', G']$ can be smaller than $G_3 = [G, G']$. In fact, we clearly have $G^{(n+1)} \leqslant G_n$ for all $n \geq 1$.

The reader will easily verify the following by induction.

LEMMA 7.4.4. *The groups G_i in the lower central series of a group G are characteristic subgroups of G.*

DEFINITION 7.4.5. A group G is *nilpotent* if its lower central series is finite.

DEFINITION 7.4.6. The *nilpotency class* of a nilpotent group is the length of its lower central series, which is to say the smallest $n \geq 0$ such that $G_{n+1} = 1$.

LEMMA 7.4.7. *Let G be a group. Then $G^{(i)} \leqslant G_i$ for all i.*

PROOF. This is almost trivial by induction, as

$$G^{(i)} = [G^{(i-1)}, G^{(i-1)}] \leqslant [G, G^{(i-1)}] \leqslant [G, G_{i-1}] = G_i.$$

□

COROLLARY 7.4.8. *Nilpotent groups are solvable.*

EXAMPLES 7.4.9.

a. The lower central series of an abelian group satisfies $G_i = 1$ for all $i \geq 1$.

b. Let T be as in Example 7.3.5. Then $T_1 = Z(T)$ and $T_2 = 1$, so T is nilpotent.

c. Let G be the group $\mathrm{Aff}(\mathbb{R})$ of upper-triangular matrices in $\mathrm{GL}_2(\mathbb{R})$ with lower-right entry 1, as in Example 2.12.13). We have

$$\left[\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}\right] = \begin{pmatrix} 1 & b(1-a) \\ 0 & 1 \end{pmatrix}$$

for all $a \in \mathbb{R}^\times$ and $b \in \mathbb{R}$. It follows easily from this that

$$G_2 = G' = \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \mid b \in \mathbb{R} \right\}$$

and then $G_i = G_2$ for all $i \geq 1$. On the other hand, G' is abelian, so $G'' = 1$. Thus, $\mathrm{Aff}(\mathbb{R})$ is solvable but not nilpotent. (Note that $\mathbb{R}$ can be replaced by any nonzero commutative ring R with unity in which $R^\times$ and $1 - R^\times$ have nontrivial intersection.)

We can give an alternative characterization of nilpotent groups through the ascending series of the following definition.

DEFINITION 7.4.10. The *upper central series* of a group G is the unique ascending series $(Z^i(G))_{i\geq 0}$ with $Z^0(G) = 1$ and $Z^{i+1}(G)$ equal to the inverse image of $Z(G/Z^i(G))$ under the projection map $G \to G/Z^i(G)$ for all $i \geq 0$.

REMARK 7.4.11. For any group G, we have $Z^1(G) = Z(G)$, and $Z^i(G) \leqslant Z^{i+1}(G)$ for all $i \geq 0$.

LEMMA 7.4.12. *If G is a nontrivial nilpotent group, then $Z(G) \neq 1$.*

PROOF. Let n be the nilpotency class of G. Then G_n is nontrivial but central in G since $[G, G_n] = 1$. □

PROPOSITION 7.4.13. *A group G is nilpotent if and only if $Z^i(G) = G$ for i sufficiently large. In this case, the nilpotency class of G equals the smallest n such that $Z^n(G) = G$, and we have $G_{n+1-i} \leqslant Z^i(G)$ for all $1 \leq i \leq n$.*

PROOF. The result is clear for abelian groupts, which are the nilpotent groups of nilpotency class 1. Let $\bar{G} = G/Z(G)$ for brevity of notation.

Suppose that G is nilpotent of nilpotency class $n \geq 2$. As $G_{n+1} = [G, G_n] = 1$, we have $1 \neq G_n \leqslant Z(G)$. Since $Z(G)$ is central in G, it follows that

$$\bar{G}_i = G_i Z(G)/Z(G)$$

for all i. Since G_{n-1} is not central in G by definition, we have that $\bar{G}_{n-1} \neq 1$. By induction on n, we then have

$$Z^{i-1}(\bar{G}) \neq \bar{G}_{n-i} \leqslant Z^i(\bar{G})$$

for $1 \leq i \leq n-1$, and $Z^{n-1}(\bar{G}) = \bar{G}$. Taking the inverse images of these groups under the quotient map from G, we obtain

$$Z^i(G) \neq G_{n-i} \leqslant Z^{i+1}(G)$$

for $1 \leq i \leq n-1$ as well, and in particular $Z^{n-1}(G) < Z^n(G) = G$.

Conversely, if $Z^n(G) = G$ for some minimial $n \geq 2$, then $Z^{n-1}(\bar{G}) = \bar{G}$, so $\bar{G}$ is nilpotent of nilpotency class $n-1$ by induction on n. This means that $G_n \leqslant Z(G)$, and therefore $G_{n+1} = 1$, so G is nilpotent. □

The following corollary can also be seen directly, using Proposition 4.9.6.

COROLLARY 7.4.14. *Finite p-groups are nilpotent.*

PROOF. This follows by induction on the order of a nontrivial p-group P, since $Z(P) \neq 1$, and $P/Z(P)$ is a p-group which we suppose by induction to be nilpotent. Then $Z^{i+1}(P)$ is the inverse image of $Z^i(P/Z(P))$ in P, so P is nilpotent as well. □

THEOREM 7.4.15 (Frattini's argument). *Let G be a finite group and N a normal subgroup. Let P be a Sylow p-subgroup of N. Then $G = N \cdot N_G(P)$.*

PROOF. For any $g \in G$, we have $gPg^{-1} \leqslant N$, as N is normal, so gPg^{-1} is also a Sylow p-subgroup of N. As such, it is conjugate to P in N, which is to say there exists $a \in N$ such that $agP(ga)^{-1} = P$, or $ag \in N_G(G)$. In other words, $g \in N \cdot N_G(P)$. □

We are now ready to prove the following equivalent conditions for nilpotency.

THEOREM 7.4.16. *Let G be a finite group. Then the following are equivalent:*

i. the group G is nilpotent,

ii. every proper subgroup of G is a proper subgroup of its normalizer in G,

iii. every Sylow p-subgroup of G is normal,

iv. G is the direct product of its Sylow p-subgroups,

v. every maximal proper subgroup of G is normal.

PROOF. Suppose that G is nilpotent of nilpotence class n, and let H be a proper subgroup of G. If $HZ(G) = G$, then H is a proper subgroup of $N_G(H) = G$. Thus, we may suppose that $HZ(G) \neq G$. As we always have that $N_G(HZ(G)) = N_G(H)$, we may further assume that $Z(G) \leqslant H$ in proving (ii). In this case, $H/Z(G)$ is a proper subgroup of $G/Z(G)$, which has nilpotence class less than n as $G_n \leqslant Z(G)$, so $(G/Z(G))_n = 1$. Thus, $H/Z(G)$ is a proper subgroup of $N_G(H/Z(G))$ by induction, but the latter group is $N_G(H)/Z(G)$ since $Z(G) \trianglelefteq N_G(H)$, and therefore $H/Z(G)$ is a proper subgroup of $N_G(H)/Z(G)$. Thus, (i) implies (ii).

Next, suppose (ii). If G is a p-group, (iii) obviously holds, so suppose this is not the case. Let P be a Sylow p-subgroup of G for some $p \mid |G|$, and note that $P < G$. Let $N = N_G(P)$. By part (ii), we have that $P < N$. Note also that P is a normal subgroup of $N_G(N)$ in that it is characteristic in N, which forces $N_G(N) = N$. Since (ii) holds, N cannot be proper in G, and thus P is normal in G. Hence, (ii) implies (iii).

Suppose (iii). Let s be the number of primes dividing $|G|$, and let $P_1, P_2, \ldots, P_s$ be the distinct Sylow subgroups of G. If $s = 1$, then we are done. In general, we set $H = P_1 \ldots P_{s-1} \lhd G$, and by induction we have that $H \cong P_1 \times P_2 \times \cdots \times P_{s-1}$. We then note that $H \cap P_s = 1$ and $HP_s = G$, so $G \cong H \times P_s$. Hence, (iii) implies (iv).

Suppose (iv), and let M be a maximal proper subgroup of G. Let $P_1, P_2, \ldots P_s$ be the distinct Sylow subgroups of G. If $M \cap P_i \neq P_i$ for some i, then $M \cap P_j = P_j$ for all $j \neq i$, since otherwise $M < MP_i < G$. Thus, M is the direct product of $M \cap P_i$ and the P_j for $j \neq i$. By the first Sylow theorem, $M \cap P_i$ is normal in P_i, so M is normal in G. Thus, (iv) implies (v).

Suppose (v). Let P be a Sylow p-subgroup of G, and suppose it is not normal. Let M be a maximal proper subgroup of G containing $N_G(P)$. Then M is normal in G, and Frattini's argument implies that $G = MN_G(P) = M$, a contradiction. So, P is normal and (iii) holds, and so (iv) holds. It then suffices to note that finite p-groups are nilpotent by Corollary 7.4.14, as this tells us that (v) implies (i). □

The following is a useful fact regarding nilpotent groups.

PROPOSITION 7.4.17. *Let G be a nilpotent group, and let S be a subset of G with image in G^{ab} a generating set. Then S generates G.*

PROOF. We prove this by induction on the nilpotence class n of nilpotent groups G. It is clear if G is abelian, or $n = 1$. For $n \geq 2$, consider G/G_n, and note that its abelianization is G^{ab}, so by induction G/G_n is generated by the image of S. Thus, if we let $H = \langle S \rangle$, we have $G = G_nH$. This implies that H is normal in G since $G_n \leqslant Z(G)$ and thus elements of G_n and of H normalize H. We have that

$$G_n = [G_nH, G_{n-1}] = [H, G_{n-1}] \leqslant H,$$

the first equality as $G_nH = G$, the second as $G_n \leqslant Z(G)$, and the third as $H \trianglelefteq G$. It follows that

$$G = G_nH = H = \langle S \rangle,$$

as claimed. □

7.5. Groups of order p^3

We note the following useful fact.

LEMMA 7.5.1. *Let G be a group such that $G/Z(G)$ is cyclic. Then G is abelian, so $G = Z(G)$.*

PROOF. Any $b \in G - Z(G)$ has image generating $G/Z(G)$, so $G = Z(G)\langle b \rangle$. As b commutes with itself and every element of $Z(G)$, it is in the center of G, a contradiction. □

Let us classify the groups of order p^3 for a prime number p.

THEOREM 7.5.2. *Let p be a prime number. There are exactly two isomorphism classes of nonabelian groups of order p^3. These are represented by:*

a. if $p = 2$, the dihedral group D_4 and the quaternion group Q_8, and

b. if p is odd, the Heisenberg group $\mathrm{Heis}(\mathbb{Z}/p\mathbb{Z})$ and the group

$$K = \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \in \mathrm{Aff}(\mathbb{Z}/p^2\mathbb{Z}) \;\middle|\; a \equiv 1 \bmod p \right\}.$$

PROOF. Let G be a nonabelian group of order p^3. By Lemma 7.5.1, the quotient $G/Z(G)$ cannot be cyclic. This eliminates the possibility that $|Z(G)| = p^2$, since then $G/Z(G)$ would be cyclic of order p. Also, $|Z(G)| \neq 1$ as G is a p-group. Thus, we have $|Z(G)| = p$, and $G/Z(G)$ is a direct product of two cyclic groups of order p. Note that $[G,G] \leqslant Z(G)$ since $G/Z(G)$ is abelian, which forces $[G,G] = Z(G)$ since G is nonabelian. Let $a, b \in G$ with images together generating $G/Z(G)$. Then $G = \langle a, b \rangle$ by Proposition 7.4.17, since finite p-groups are nilpotent. Moreover, $z = [b,a]$ generates $Z(G)$, and note that this means $ab = baz$.

Now, suppose that G has an element of order p^2. Without loss of generality, we may suppose that it is b. Then b^p generates $Z(G)$, so we have $b^{pi} = z$ for some $i \in \mathbb{Z}$ with $0 < i < p$. We then have $ab = b^{1+pi}a$, which in particular tells us that $\langle b \rangle$ is a normal subgroup of G. Suppose that we can also choose a to have order p. Then $G = \langle b \rangle \rtimes \langle a \rangle$, and in fact G has a presentation

$$G \cong \langle a, b \mid a^p = b^{p^2} = e, ab = b^{1+pi}a \rangle.$$

If p is odd, then we have an isomorphism

$$f\colon G \to K, \qquad f(a) = \begin{pmatrix} 1+pi & 0 \\ 0 & 1 \end{pmatrix},\ f(b) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

and if $p = 2$, then $f\colon G \to D_4$ defined by $f(a) = s$, $f(b) = r$ works as well.

Suppose now that we cannot choose a and b with $G = \langle a,b\rangle$ and either a or b of order p. If $p = 2$, then this implies that $a^2 = b^2 = (ab)^2 = z$, and from this one checks that G is a quotient of the group with presentation

$$\langle x,y \mid x^2 = y^2 = (xy)^2, x^4 = e\rangle,$$

and the latter group is isomorphic to Q_8 under the map which takes x to i and y to j. If p is odd, then one sees that

$$(ba)^p = b^p a^p z^{p(p-1)/2} = b^p a^p = z^{pi} z^{pj} = z^{p(i+j)}$$

for some $i, j \in \mathbb{Z}$ prime to p. Note that we have used p is odd here, since otherwise $p(p-1)/2 = 1$, which is not a multiple of $p = 2$. If we replace a by a^k where $k \in \mathbb{Z}$ with $ik \equiv -j \bmod p$, then $(ba)^p = e$, which yields a contradiction.

Finally, suppose that G has no element of order p^2. If $p = 2$, then $z = [b,a] = (ba)^2 = e$, which is a contradiction. If p is odd, then G is a quotient of the group that has presentation

$$\langle x,y \mid x^p, y^p, [x,y]^p, [x,[x,y]], [y,[x,y]]\rangle,$$

but $\mathrm{Heis}(\mathbb{Z}/p\mathbb{Z})$ has this presentation, so it is isomorphic to G in that G has order p^3. □

CHAPTER 8

Category theory

8.1. Categories

The extremely broad concept of a "category" allows us to deal with many of the constructions in mathematics in an abstract context. We begin with the definition. We will mostly ignore set-theoretical considerations that can be used to put what follows on a firmer basis, but note that a *class* is a collection of objects that can be larger than a set, e.g., the class of all sets, in order that we might avoid Russell's paradox.

DEFINITION 8.1.1. A *category* $\mathscr{C}$ is

(1) a class of *objects* $\mathrm{Obj}(\mathscr{C})$,

(2) for every $A,B \in \mathrm{Obj}(\mathscr{C})$, a class $\mathrm{Hom}_{\mathscr{C}}(A,B)$ of *morphisms* from A to B, where we often use the notation $f\colon A \to B$ to indicate that f is an element of $\mathrm{Hom}_{\mathscr{C}}(A,B)$, and

(3) a *composition* map

$$\mathrm{Hom}_{\mathscr{C}}(A,B) \times \mathrm{Hom}_{\mathscr{C}}(B,C) \to \mathrm{Hom}_{\mathscr{C}}(A,C)$$

for each $A,B,C \in \mathrm{Obj}(\mathscr{C})$ that takes (f,g) for $f\colon A \to B$ and $g\colon B \to C$ to the *composition* $g \circ f$, subject to the properties that

i. for each $A \in \mathrm{Obj}(\mathscr{C})$, there exists an *identity morphism* $\mathrm{id}_A\colon A \to A$ such that, for all $f\colon A \to B$ and $g\colon B \to A$ with $B \in \mathrm{Obj}(\mathscr{C})$, we have

$$f \circ \mathrm{id}_A = f \quad \text{and} \quad \mathrm{id}_A \circ g = g,$$

and

ii. composition is *associative*, i.e.,

$$h \circ (g \circ f) = (h \circ g) \circ f$$

for any three morphisms $h\colon C \to D$, $g\colon B \to C$, and $f\colon A \to B$ between objects $A,B,C,D \in \mathrm{Obj}(\mathscr{C})$.

DEFINITION 8.1.2. We say that a category is *small* if its class of objects is a set.

REMARK 8.1.3. What we call a category is often referred to as a *locally category*, and a *category* in that terminology allows the morphisms between a pair of objects to form a class, not just a set.

EXAMPLES 8.1.4.

a. The category **Set** which has sets as its objects and maps of sets as its morphisms.

b. The category **Gp** which has groups as its objects and group homomorphisms as it morphisms.

c. Similarly, we have categories **Ring**, the objects of which we take to be the (possibly zero) rings with 1 and with morphisms the ring homomorphisms that preserve 1, and **Field**.

d. If R is a ring, then the category R-**mod** has objects the left R-modules and morphisms the left R-module homomorphisms.

e. The category **Top** which has topological spaces as its objects and continuous maps as its morphisms.

We may construct new categories out of old. The following provides a useful example.

DEFINITION 8.1.5. Let $\mathscr{C}$ and $\mathscr{D}$ be categories. The *product category* $\mathscr{C} \times \mathscr{D}$ is the category with objects the pairs (C,D) with $C \in \mathrm{Obj}(\mathscr{C})$ and $D \in \mathrm{Obj}(\mathscr{D})$ and morphisms $(f,g)\colon (C,D) \to (C',D')$ for any $f\colon C \to C'$ in $\mathscr{C}$ and $g\colon D \to D'$ in $\mathscr{D}$.

DEFINITION 8.1.6. Given a category $\mathscr{C}$, we define the *opposite category* $\mathscr{C}^{\mathrm{op}}$ to have the same class of objects as $\mathscr{C}$ and

$$\mathrm{Hom}_{\mathscr{C}^{\mathrm{op}}}(A,B) = \mathrm{Hom}_{\mathscr{C}}(B,A)$$

for $A,B \in \mathrm{Obj}(\mathscr{C})$.

DEFINITION 8.1.7. A *monoid* G is a set with an associative binary operation and an identity element for the operation.

EXAMPLE 8.1.8. Any monoid G gives rise to a category with one object, morphisms equal to the elements of G, and composition law given by multiplication. Then G^{op} is again a monoid with the same elements but the multiplication reversed. A category with one object is also called a *monoid*, and we have a one-to-one correspondence between monoids and these categories.

We will often have cause to single out a particular class of morphisms in a category known as isomorphisms.

DEFINITION 8.1.9. Let $\mathscr{C}$ be a category.

a. A morphism $f\colon A \to B$ in $\mathscr{C}$ is an *isomorphism* if there exists morphism $g\colon B \to A$ in $\mathscr{C}$ such that $g \circ f = \mathrm{id}_A$ and $f \circ g = \mathrm{id}_B$.

b. Two objects A and B in $\mathscr{C}$ are said to be *isomorphic* if there exists an isomorphism $f\colon A \to B$ in $\mathscr{C}$.

c. If $f\colon A \to B$ is a morphism and $g \circ f = \mathrm{id}_A$ (resp., $f \circ g = \mathrm{id}_B$), then we say that g is a *right inverse* (resp., a *left inverse*) to f. If both $g \circ f = \mathrm{id}_A$ and $f \circ g = \mathrm{id}_B$, then we say that g is (an) *inverse* to f, or that f and g are *inverse to each other* (or *mutually inverse*, or *inverses*).

EXAMPLES 8.1.10.

a. The isomorphisms in **Set** are the bijections.

b. The isomorphisms in **Gp** are the isomorphisms of groups.

c. The isomorphisms in **Top** are the homeomorphisms.

DEFINITION 8.1.11.

a. A morphism $f\colon A \to B$ in a category $\mathscr{C}$ is a *monomorphism* if for any $g,h\colon C \to A$ with $C \in \mathrm{Obj}(\mathscr{C})$, the property that $f \circ g = f \circ h$ implies $g = h$.

b. A morphism $f\colon A \to B$ in a category $\mathscr{C}$ is an *epimorphism* if for any $g,h\colon B \to C$ with $C \in \mathrm{Obj}(\mathscr{C})$, the property that $g \circ f = h \circ f$ implies $g = h$.

EXAMPLES 8.1.12.

a. In **Set** and R-**mod**, a morphism is a monomorphism (resp., epimorphism) if and only if it is injective (resp., surjective).

b. The natural injection $\mathbb{Z} \to \mathbb{Q}$ in **Ring** is an epimorphism, since a ring homomorphism $\mathbb{Q} \to R$ is completely determined by its value on 1.

REMARK 8.1.13. A morphism $f\colon A \to B$ in a category $\mathscr{C}$ is a monomorphism if and only if the opposite morphism $f^{\mathrm{op}}\colon B \to A$ in $\mathscr{C}^{\mathrm{op}}$ is an epimorphism.

We have the following.

LEMMA 8.1.14. *Let $f\colon A \to B$ and $g\colon B \to A$ be morphisms in a category $\mathscr{C}$ such that $g \circ f = \mathrm{id}_A$. Then f is a monomorphism and g is an epimorphism.*

PROOF. Let $h,k\colon C \to A$ be morphisms such that $f \circ h = f \circ k$. Then

$$k = g \circ f \circ k = g \circ f \circ h = h.$$

Thus f is a monomorphism. Similarly, g is an epimorphism, or apply Remark 8.1.13. □

In other words, right inverses are monomorphisms and left inverses are epimorphisms.

DEFINITION 8.1.15. Let $\mathscr{C}$ be a category and $C \in \mathrm{Obj}(\mathscr{C})$.

a. A *subobject* of C is a pair (A,ι) consisting of an object A and a monomorphism $\iota\colon A \to C$.

b. A *quotient* of C is a pair (B,π) consisting of an object B and an epimorphism $\pi\colon C \to B$.

DEFINITION 8.1.16. A *subcategory* $\mathscr{C}$ of a category $\mathscr{D}$ is a category with objects consisting of a subclass of $\mathrm{Obj}(\mathscr{D})$ and morphisms $\mathrm{Hom}_{\mathscr{C}}(A,B)$ for $A,B \in \mathrm{Obj}(\mathscr{C})$ consisting of a subset of $\mathrm{Hom}_{\mathscr{D}}(A,B)$ containing id_A for $A = B$ and such that composition maps in $\mathscr{C}$ agree with the restriction of the composition maps in $\mathscr{D}$ between the same objects.

EXAMPLES 8.1.17.

a. The category **Ab** of abelian groups with morphisms the group homomorphisms between abelian groups is a subcategory of **Gp**.

b. The category **Field** is a subcategory of **Ring**.

8.2. Functors

To compare two categories, we need some notion of a map between them. Such maps are referred to as functors. There are two basic types.

DEFINITION 8.2.1. Let $\mathscr{C}$ and $\mathscr{D}$ be categories.

a. A *covariant functor* (or simply *functor*) $F\colon \mathscr{C} \to \mathscr{D}$ between two categories $\mathscr{C}$ and $\mathscr{D}$ is a map of objects $F\colon \operatorname{Obj}(\mathscr{C}) \to \operatorname{Obj}(\mathscr{D})$ and a map of morphisms

$$F\colon \operatorname{Hom}_{\mathscr{C}}(A,B) \to \operatorname{Hom}_{\mathscr{D}}(F(A),F(B))$$

for each $A,B \in \operatorname{Obj}(\mathscr{C})$ such that $F(\mathrm{id}_A) = \mathrm{id}_{F(A)}$ and $F(g \circ f) = F(g) \circ F(f)$ for all $f\colon A \to B$ and $g\colon B \to C$ for each $A,B,C \in \operatorname{Obj}(\mathscr{C})$.

b. As with a covariant functor, a *contravariant functor* $F\colon \mathscr{C} \to \mathscr{D}$ is again a map on objects, but with maps between sets of morphisms of the form

$$F\colon \operatorname{Hom}_{\mathscr{C}}(A,B) \to \operatorname{Hom}_{\mathscr{D}}(F(B),F(A))$$

that satisfies $F(\mathrm{id}_A) = \mathrm{id}_{F(A)}$ and $F(g \circ f) = F(f) \circ F(g)$.

We give some examples of functors.

EXAMPLES 8.2.2.

a. We have the *forgetful functors* **Gp** → **Set**, **Ring** → **Set**, and **Top** → **Set**, which take objects to their underlying sets and morphisms to the corresponding set-theoretic maps.

b. We have another forgetful functor from R-**mod** to the category **Ab** of abelian groups.

c. A homomorphism of monoids $G \to G'$ induces a functor of the corresponding categories, and conversely.

d. The opposite functor $\mathrm{op}\colon \mathscr{C} \to \mathscr{C}^{\mathrm{op}}$ that is the identity on objects and takes a morphism $f\colon A \to B$ to its opposite morphism $f^{\mathrm{op}}\colon B \to A$ in $\mathscr{C}^{\mathrm{op}}$ is contravariant.

REMARK 8.2.3. A contravariant functor $F\colon \mathscr{C} \to \mathscr{D}$ may also be viewed as a covariant functor $\mathscr{C} \to \mathscr{D}^{\mathrm{op}}$, in particular by composing F with the opposite functor $\mathrm{op}\colon \mathscr{D} \to \mathscr{D}^{\mathrm{op}}$.

REMARK 8.2.4. A subcategory $\mathscr{C}$ of a category $\mathscr{D}$ is endowed with a canonical inclusion functor that takes an object of $\mathscr{C}$ to the same object of $\mathscr{D}$ and is the identity map on morphism.

DEFINITION 8.2.5. Let $F\colon \mathscr{C} \to \mathscr{D}$ be a functor.

a. The functor F is called *faithful* if it is one-to-one on morphisms.

b. The functor F is called *full* if it is onto on morphisms.

c. A functor F is *fully faithful* if it is both faithful and full.

d. A subcategory is called a *full subcategory* if it the corresponding inclusion functor is full.

REMARK 8.2.6. Every functor takes isomorphisms to isomorphisms.

REMARK 8.2.7. The inclusion functor attached to a subcategory is always faithful.

REMARK 8.2.8. A fully faithful functor is sometimes referred to as an embedding of categories, or sometimes a full embedding (and when so, a faithful but not necessarily full functor might instead be referred to as an embedding).

EXAMPLES 8.2.9.

a. The category **Ab** is a full subcategory of **Gp**.

b. The category **Field** is a full subcategory of **Ring**.

c. The above-described forgetful functors to sets are faithful but not full.

d. The category in which the objects are sets but the morphisms are bijections of sets is a subcategory of **Set** that has the same objects but is not full.

DEFINITION 8.2.10. A *directed graph* $\mathscr{G}$ is a collection consisting of

(1) a set $V_{\mathscr{G}}$ of *vertices* of $\mathscr{G}$ and,

(2) for every $v, w \in V_{\mathscr{G}}$, a set $E_{\mathscr{G}}(v, w)$ of *edges* from v to w in $\mathscr{G}$.

TERMINOLOGY 8.2.11. In category theory, we often refer to the vertices of a directed graph as *dots* and the edges as *arrows*.

EXAMPLE 8.2.12. The following picture provides the data of a directed graph with 4 vertices and edge sets with between 0 and 2 elements each:

$$\begin{tikzcd}
\cdot \arrow[r] \arrow[r, bend left] & \cdot \arrow[d] \\
\cdot \arrow[r] & \cdot \arrow[u, bend right]
\end{tikzcd}$$

DEFINITION 8.2.13. The *category (freely) generated by a directed graph* $\mathscr{G}$ is the category I with $\mathrm{Obj}(I) = V_{\mathscr{G}}$ and, for $v, w \in \mathrm{Obj}(I)$, with $\mathrm{Hom}_I(v, w)$ equal to the set of all words $e_n e_{n-1} \cdots e_1$ for some $n \geq 0$ (with $n = 0$ providing the empty word) with $e_i \in E_{\mathscr{G}}(v_{i-1}, v_i)$ for $v_i \in V_{\mathscr{G}}$ for $1 \leq i \leq n$, with $v_0 = v$ and $v_n = w$, together with the composition given by concatenation of words.

EXAMPLE 8.2.14. Consider the directed graph $\mathscr{G}$ given by

$$\begin{tikzcd}
v_1 \arrow[r, "e_1"] & v_2 \arrow[r, "e_2"] & v_3.
\end{tikzcd}$$

The category I generated by $\mathscr{G}$ has three objects v_1, v_2, v_3 and morphism sets

$$\begin{aligned} \mathrm{Hom}_I(v_i, v_i) &= \{\mathrm{id}_{v_i}\}, & \mathrm{Hom}_I(v_i, v_{i+1}) &= \{e_i\}, \\ \mathrm{Hom}_I(v_1, v_3) &= \{e_2 e_1\}, \quad \text{and} & \mathrm{Hom}_I(v_i, v_j) &= \varnothing \text{ if } j < i. \end{aligned}$$

EXAMPLE 8.2.15. Consider the directed graph $\mathscr{G}$ given by

$$\begin{tikzcd}
v_1 \arrow[r, bend left, "e_1"] & v_2. \arrow[l, bend left, "e_2"]
\end{tikzcd}$$

Let I be the category generated $\mathscr{G}$. For $i, j \in \{1, 2\}$ the set $\mathrm{Hom}_I(v_i, v_j)$ consists of the words with alternating letters e_1 and e_2 that start with e_i and end with e_j (including the empty word if $i = j$).

DEFINITION 8.2.16. A *diagram* in $\mathscr{C}$ is a functor from a category generated by a graph to $\mathscr{C}$.

REMARK 8.2.17. Let $\mathscr{G}$ be a directed graph, let I be the category generated by $\mathscr{G}$, and let $\mathscr{C}$ be a category. Given a map $F\colon V_{\mathscr{G}} \to \mathscr{C}$ and functions $F\colon E_{\mathscr{G}}(v,w) \to \mathrm{Hom}_{\mathscr{C}}(F(v),F(w))$ for each $v,w, \in V_{\mathscr{G}}$, there exists a unique functor $F\colon I \to \mathscr{C}$ that agrees with F on $V_{\mathscr{G}}$ and on $E_{\mathscr{G}}(v,w) \subseteq \mathrm{Hom}_I(v,w)$ for every $v,w \in V_{\mathscr{G}}$.

REMARK 8.2.18. Often, we consider finite graphs, in which every collection of vertices and edges is finite. The resulting diagrams are known as finite diagrams.

DEFINITION 8.2.19. A *commutative diagram* in $\mathscr{C}$ is a diagram $F\colon I \to \mathscr{C}$, where I is the category generated by a graph, which is a constant function on every set of morphisms.

EXAMPLE 8.2.20. To give a functor from I as in Example 8.2.14 to a category $\mathscr{C}$ is to proscribe three objects A,B,C in $\mathscr{C}$ and two morphisms $f\colon A \to B$ and $g\colon B \to C$. Thus, such a diagram may be represented by

$$A \xrightarrow{f} B \xrightarrow{g} C,$$

and it is automatically commutative.

EXAMPLE 8.2.21. To give a functor from I as in Example 8.2.15 to a category $\mathscr{C}$ is to proscribe two objects A,B in $\mathscr{C}$ and two morphisms $f\colon A \to B$ and $g\colon B \to A$. The diagram

$$\begin{tikzcd} A \arrow[r, bend left, "f"] & B \arrow[l, bend left, "g"] \end{tikzcd}$$

is commutative if and only if $f \circ g = \mathrm{id}_B$ and $g \circ f = \mathrm{id}_A$.

8.3. Natural transformations

DEFINITION 8.3.1. Let $F,G\colon \mathscr{C} \to \mathscr{D}$ be two (covariant) functors. A *natural transformation* $\eta\colon F \rightsquigarrow G$ is a class of morphisms $\eta_A\colon F(A) \to G(A)$ for each $A \in \mathrm{Obj}(\mathscr{C})$ subject to the condition that

$$\begin{tikzcd} F(A) \arrow[r, "\eta_A"] \arrow[d, "F(f)"'] & G(A) \arrow[d, "G(f)"] \\ F(B) \arrow[r, "\eta_B"] & G(B) \end{tikzcd}$$

commutes for every $f\colon A \to B$ and $A,B \in \mathrm{Obj}(\mathscr{C})$. If instead F and G are contravariant functors, then the direction of the vertical arrows in the diagram are reversed.

EXAMPLE 8.3.2. Consider the functor $T\colon \mathbf{Ab} \to \mathbf{Ab}$ that sends an abelian group A to its torsion subgroup A_{tor} (i.e., the subgroup of elements of finite order) and takes a homomorphism $f\colon A \to B$ to its restriction $T(f)\colon A_{\mathrm{tor}} \to B_{\mathrm{tor}}$. Let $I\colon \mathbf{Ab} \to \mathbf{Ab}$ denote the identity functor. For each abelian group A, we can define $\iota_A\colon A_{\mathrm{tor}} \to A$ to be the inclusion map. We clearly have $f \circ \iota_A = \iota_B \circ T(f)$ for all $f\colon A \to B$, so $\iota\colon T \rightsquigarrow I$ is a natural transformation.

EXAMPLE 8.3.3. If we think of groups G and G' as monoids, so that functors $G \to G'$ are homomorphisms, then a natural transformation $f \rightsquigarrow f'$ between two homomorphisms $f,f'\colon G \to G'$ is given simply by an element $x \in G'$ such that $f'(g) = xf(g)x^{-1}$ for all $g \in G$.

DEFINITION 8.3.4. Let $F,G\colon \mathscr{C} \to \mathscr{D}$ be functors. A natural transformation $\eta\colon F \rightsquigarrow G$ is said to be a natural isomorphism if each η_A for $A \in \mathscr{C}$ is an isomorphism.

REMARK 8.3.5. Every natural isomorphism $\eta\colon F \rightsquigarrow G$ has an inverse $\eta^{-1}\colon G \rightsquigarrow F$ with $\eta_A^{-1} = (\eta_A)^{-1}$ for $A \in \mathrm{Obj}(\mathscr{C})$.

DEFINITION 8.3.6. Let $F,G\colon \mathscr{C} \to \mathscr{D}$ be functors, and let $\eta_A\colon F(A) \to G(A)$ be morphisms for each $A \in \mathrm{Obj}(\mathscr{C})$. We say that these morphisms are *natural* if the η_A form a natural transformation $\eta\colon F \to G$.

DEFINITION 8.3.7. Two categories $\mathscr{C}$ and $\mathscr{D}$ are said to be *equivalent* if there exist functors $F\colon \mathscr{C} \to \mathscr{D}$ and $G\colon \mathscr{D} \to \mathscr{C}$ and natural isomorphisms $\eta\colon G \circ F \rightsquigarrow \mathrm{id}_{\mathscr{C}}$ and $\eta'\colon F \circ G \rightsquigarrow \mathrm{id}_{\mathscr{D}}$. Two such functors F and G are said to be *quasi-inverse*, and F and G are said to be *equivalences of categories*.

EXAMPLE 8.3.8. A category $\mathscr{C}$ with one object 0 and one morphism is equivalent to the category $\mathscr{D}$ with two objects 1, 2 and four morphisms, the identity morphisms of 1 and 2 and isomorphisms $1 \to 2$ and $2 \to 1$. We have quasi-inverse functors F and G with $F(0) = 1$ and $F(\mathrm{id}_0) = \mathrm{id}_1$ and $G(1) = G(2) = 0$ and $G(f) = \mathrm{id}_0$ for all f. To see naturality, note that every morphism between two objects in either category is unique.

The following theorem provides a standard example of equivalence of categories.

THEOREM 8.3.9 (Morita equivalence). *The category of left modules over a ring R with unity is equivalent to the category of left modules over $M_n(R)$ for every $n \geq 1$.*

PROOF. Let A be the R-$M_n(R)$-bimodule of row vectors of length n with R-entries. Let B be the $M_n(R)$-R-bimodule of column vectors of length n with R-entries. Define

$$F\colon R\text{-}\mathbf{mod} \to M_n(R)\text{-}\mathbf{mod}, \quad F(M) = B \otimes_R M, \quad F(f) = \mathrm{id}_B \otimes f$$

for left R-modules M and M' and $f \in \mathrm{Hom}_R(M,M')$. Also, define

$$G\colon M_n(R)\text{-}\mathbf{mod} \to R\text{-}\mathbf{mod}, \quad G(N) = A \otimes_{M_n(R)} N, \quad G(g) = \mathrm{id}_A \otimes g$$

for left $M_n(R)$-modules N and N' and $g \in \mathrm{Hom}_{M_n(R)}(N,N')$. Since multiplication induces isomorphisms

$$A \otimes_{M_n(R)} B \xrightarrow{\sim} R \quad \text{and} \quad B \otimes_R A \xrightarrow{\sim} M_n(R),$$

both $G \circ F$ and $F \circ G$ are naturally isomorphic to identity functors. □

DEFINITION 8.3.10. Given two categories $\mathscr{C}$ and $\mathscr{D}$ with $\mathscr{C}$ small, the *functor category* $\mathbf{Func}(\mathscr{C},\mathscr{D})$ has objects the functors $\mathscr{C} \to \mathscr{D}$ and morphisms the natural transformations between functors, defining composition of natural transformations via composition of the morphisms determining them.

DEFINITION 8.3.11. Let $\mathscr{C}$ be a category and $A \in \mathrm{Obj}(\mathscr{C})$ be an object.

a. We have a functor $h_A\colon \mathscr{C} \to \mathrm{Sets}$ given by

$$h_A(B) = \mathrm{Hom}_{\mathscr{C}}(A,B)$$

and, for $g\colon B \to C$,

$$h_A(g)(f) = g \circ f$$

for all $f\colon A \to B$.

b. We have a contravariant functor $h^A\colon \mathscr{C} \to \mathbf{Set}$ with

$$h^A(B) = \mathrm{Hom}_{\mathscr{C}}(B,A) \quad \text{and} \quad h^A(g)(f) = f \circ g$$

for $B, C \in \mathrm{Obj}(\mathscr{C})$, $g\colon B \to C$, and $f\colon C \to A$.

DEFINITION 8.3.12. Let $\mathscr{C}$ be a small category. The *Yoneda embedding* is the functor

$$h^{\mathscr{C}}\colon \mathscr{C} \to \mathbf{Func}(\mathscr{C}^{\mathrm{op}}, \mathbf{Set})$$

defined by $h^{\mathscr{C}}(A) = h^A$ for $A \in \mathrm{Obj}(\mathscr{C})$ and $h^{\mathscr{C}}(f)\colon h^A \rightsquigarrow h^B$ for $f\colon A \to B$ in $\mathscr{C}$ given by

$$h^{\mathscr{C}}(f)_C(g) = f \circ g$$

for each $g\colon C \to A$ in $\mathscr{C}$ and any $C \in \mathrm{Obj}(\mathscr{C})$.

REMARK 8.3.13. The reader should check that the Yoneda embedding is a well-defined functor.

THEOREM 8.3.14. *Let $\mathscr{C}$ be a small category. The Yoneda embedding $h^{\mathscr{C}}$ is fully faithful.*

PROOF. We first show faithfulness. Let $f, g\colon A \to B$ be two morphisms with $h^{\mathscr{C}}(f) = h^{\mathscr{C}}(g)$. Then

$$f = f \circ \mathrm{id}_A = h^{\mathscr{C}}(f)_A(\mathrm{id}_A) = h^{\mathscr{C}}(g)_A(\mathrm{id}_A) = g \circ \mathrm{id}_A = g.$$

As for fullness, suppose that $\eta\colon h^A \rightsquigarrow h^B$ for some $A, B \in \mathrm{Obj}(\mathscr{C})$. We claim that $\eta = h(e)$, where $e = \eta_A(\mathrm{id}_A)$. To see this, note that the fact that η is a natural transformation means, in particular, that the diagram

$$\begin{tikzcd}
h^A(A) \arrow[r, "\eta_A"] \arrow[d, "h^A(f)"'] & h^B(A) \arrow[d, "h^B(f)"] \\
h^A(C) \arrow[r, "\eta_C"] & h^B(C)
\end{tikzcd}$$

commutes for any $f\colon C \to A$. Applying both compositions to the identity morphism of A, we get the two equal terms

$$h^B(f) \circ \eta_A(\mathrm{id}_A) = h^B(f)(e) = f \circ e = h(e)_C(f)$$

and

$$\eta_C \circ h^A(f)(\mathrm{id}_A) = \eta_C(\mathrm{id}_A \circ f) = \eta_C(f),$$

and therefore, the desired equality. □

REMARK 8.3.15. Similarly, we have a fully faithful contravariant functor

$$h_{\mathscr{C}}\colon \mathscr{C} \to \mathbf{Func}(\mathscr{C}, \mathbf{Set})$$

given by the h_A for $A \in \mathrm{Obj}(\mathscr{C})$ and natural transformations between them. This is just the Yoneda embedding for the category $\mathscr{C}^{\mathrm{op}}$.

Theorem 8.3.14 can be thought of as a more general version of the following standard theorem of group theory.

COROLLARY 8.3.16 (Cayley's theorem). *Every group G is isomorphic to a subgroup of the symmetric group S_G on G.*

PROOF. Consider the monoid $\mathbb{G}$ formed by G. Recall that in $\mathbb{G}$, morphisms are elements of G. As $h\colon \mathbb{G} \to \mathbf{Func}(\mathbb{G}^{\mathrm{op}}, \mathbf{Set})$ is a functor, Yoneda's lemma provides an injective function

$$h\colon G \to \mathrm{Hom}_{\mathbf{Func}(\mathbb{G}^{\mathrm{op}},\mathbf{Set})}(h^G, h^G)$$

on morphisms with the properties that $h(e) = \mathrm{id}_{h^G}$ and $h(xy) = h(x) \circ h(y)$ for $x, y \in G$. Since $\mathbb{G}$ has only the object G, and $h^G(G) = G$, this induces a one-to-one function $\rho\colon G \to \mathrm{Maps}(G,G)$ with $\rho(x) = h(x)_G$ and satisfying $\rho(xy) = \rho(x) \circ \rho(y)$ and $\rho(e) = \mathrm{id}_G$. In particular, we have $\rho(x^{-1}) \circ \rho(x) = \mathrm{id}_G$ for every $x \in G$, so its image lands in S_G, and the resulting map $G \to S_G$ is an injective homomorphism. □

We shall later require the following strengthening of Theorem 8.3.14.

THEOREM 8.3.17 (Yoneda's lemma). *For any object A of a small category $\mathscr{C}$ and contravariant functor $F\colon \mathscr{C} \to \mathbf{Set}$, there is a bijection*

$$\mathrm{Hom}_{\mathbf{Func}(\mathscr{C}^{\mathrm{op}},\mathbf{Set})}(h^A, F) \xrightarrow{\sim} F(A)$$

given by $\eta \mapsto \eta_A(\mathrm{id}_A)$ that is natural in A and F.

PROOF. Let $B \in \mathrm{Obj}(\mathscr{C})$. Given $x \in F(A)$, consider the composition

$$\mathrm{Hom}_{\mathscr{C}}(B,A) \xrightarrow{F} \mathrm{Hom}_{\mathbf{Set}}(F(A), F(B)) \xrightarrow{\mathrm{ev}_x} F(A),$$

where ev_x is evaluation at x. This defines a natural transformation $\xi^x\colon h^A \rightsquigarrow F$. If $\eta\colon h^A \rightsquigarrow F$ and $f\colon B \to A$, then

$$F(f) \circ \eta_A(\mathrm{id}_A) = \eta_B(\mathrm{id}_A \circ f) = \eta_B(f)$$

by the naturality of η. On the other hand, if $x \in F(A)$, then

$$\xi^x_A(\mathrm{id}_A) = \mathrm{ev}_x(F(\mathrm{id}_A)) = \mathrm{ev}_x(\mathrm{id}_{F(A)}) = x.$$

Hence the maps $\eta \mapsto \eta_A(\mathrm{id}_A)$ and $x \mapsto \xi^x$ are inverse to each other. □

8.4. Limits and colimits

In this section, $\mathscr{C}$ denotes a category, and I denotes a small category.

NOTATION 8.4.1. We write $i \in I$ to denote, more simply, that i is an object in I.

DEFINITION 8.4.2. Let $F\colon I \to \mathscr{C}$ be a functor. When it exists, the *limit* of F is a pair $(\lim F, (\phi_i)_{i \in I})$ consisting of an object $\lim F$ in $\mathscr{C}$ and morphisms

$$\phi_i\colon \lim F \to F(i)$$

for each $i \in I$ such that $\phi_j = F(\kappa) \circ \phi_i$ for all morphisms $\kappa\colon i \to j$ in I and with the *universal property* that if X is any object of $\mathscr{C}$ together with morphisms $\psi_i\colon X \to F(i)$ for which $\psi_j =$

$\kappa \circ \psi_i$ for all morphisms $\kappa\colon i \to j$, then there exists a unique morphism $f\colon X \to \lim F$ such that $\psi_i = \phi_i \circ f$ for all $i \in I$.

NOTATION 8.4.3. We usually use $\lim F$ to refer more simply to a pair $(\lim F, (\phi_i)_{i\in I})$ that is a limit of $F\colon I \to \mathscr{C}$, with the maps understood.

REMARK 8.4.4. The universal property of the limit of a functor F as in Definition 8.4.2 may be visualized by commutative diagrams

$$\begin{tikzcd}
 & X \arrow[d, "f"] \arrow[ddl, bend right, "\psi_i"'] \arrow[ddr, bend left, "\psi_j"] & \\
 & \lim F \arrow[dl, "\phi_i"'] \arrow[dr, "\phi_j"] & \\
F(i) \arrow[rr, "F(\kappa)"] & & F(j).
\end{tikzcd}$$

LEMMA 8.4.5. *If $(X, (\psi_i)_{i\in I})$ and $(\lim F, (\phi_i)_{i\in I})$ are limits of a functor $F\colon I \to \mathscr{C}$, then there is a unique isomorphism $f\colon X \to \lim F$ such that $\psi_i = \phi_i \circ f$ for all $i \in I$.*

PROOF. There are morphisms $f\colon X \to \lim F$ and $g\colon \lim F \to X$ that are unique with the respective properties that $\psi_i = \phi_i \circ f$ and $\phi_i = \psi_i \circ g$ for all $i \in I$. Note that we have $\phi_i = \phi_i \circ f \circ g$ for all $i \in I$. On the other hand, the universal property of X implies that the identity id_X is the unique morphism h such that $\psi_i \circ h = \psi_i$ for all $i \in I$, so $f \circ g = \mathrm{id}_X$. Similarly, $g \circ f$ is the identity of $\lim F$ by its universal property. Therefore, the unique map f is an isomorphism. □

REMARK 8.4.6. Lemma 8.4.5 says that a limit, when it exists, is unique up to unique isomorphism (respecting the universal property) and for that reason, we refer to "the", rather than "a", limit.

If I has only identity morphisms, then the limit of a functor $F\colon I \to \mathscr{C}$ is determined entirely by the image objects $A_i = F(i)$ for all $i \in I$. Hence the notation in the following definition makes sense.

DEFINITION 8.4.7. Let I be a category with only identity morphisms, and let $F\colon I \to \mathscr{C}$ be a functor. Set $A_i = F(i)$ for each $i \in I$.

a. The limit $\prod_{i\in I} A_i$ of F, when it exists, is called the *product* of the A_i.

b. The maps
$$p_i\colon \prod_{i\in I} A_i \to A_i$$
resulting from the universal property of the product are known as *projection maps*.

EXAMPLES 8.4.8. The product coincides with direct product in the categories **Set**, **Gp**, **Top**, **Ring** and R-**mod**. Products of more than one object do not exist in the category **Field**.

REMARK 8.4.9. A commutative diagram in a category $\mathscr{C}$ arises from a functor $F\colon I \to \mathscr{C}$, where I is a category generated by a directed graph. Therefore, we may speak of the limit of the diagram.

DEFINITION 8.4.10. The limit $A_1 \times_B A_2$ of a diagram

$$\begin{tikzcd} & A_1 \arrow[d, "f_1"] \\ A_2 \arrow[r, "f_2"] & B \end{tikzcd} \tag{8.4.1}$$

in $\mathscr{C}$, when it exists, is called the *pullback* of the diagram.

REMARK 8.4.11. The pullback of (8.4.1) is endowed with morphisms p_1 and p_2 that make

$$\begin{tikzcd} A_1 \times_B A_2 \arrow[r, "p_1"] \arrow[d, "p_2"'] & A_1 \arrow[d, "f_1"] \\ A_2 \arrow[r, "f_2"] & B \end{tikzcd}$$

commute.

EXAMPLE 8.4.12. In **Set**, **Gp**, **Top**, and *R*-**mod**, the pullback is the subobject (i.e., subset, subgroup, subspace, or submodule) with underlying set

$$\{(a_1, a_2) \in A_1 \times A_2 \mid f_1(a_1) = f_2(a_2)\}.$$

We also have the dual notion to limits:

DEFINITION 8.4.13. Let $F\colon I \to \mathscr{C}$ be a functor. When it exists, the *colimit* of F is a pair $(\operatorname{colim} F, (\alpha_i)_{i \in I})$ consisting of an object $\operatorname{colim} F \in \mathscr{C}$ together with morphisms

$$\alpha_i\colon F(i) \to \operatorname{colim} F$$

for each $i \in I$ such that $\alpha_j \circ F(\kappa) = \alpha_i$ for all morphisms $\kappa\colon i \to j$ and with the *universal property* that if X is any object of $\mathscr{C}$ together with morphisms $\beta_i\colon X \to F(i)$ for which $\beta_j \circ \kappa = \beta_i$ for all morphisms $\kappa\colon i \to j$, then there exists a unique morphism $f\colon \operatorname{colim} F \to X$ such that $\beta_i = f \circ \alpha_i$ for all $i \in I$.

NOTATION 8.4.14. A colimit of a functor $F\colon I \to \mathscr{C}$ is usually denoted simply by the object $\operatorname{colim} F$, with the morphisms omitted.

REMARK 8.4.15. The properties of the colimit expressed in Definition 8.4.13 may be summarized by the commutativity of the diagrams

$$\begin{tikzcd} F(i) \arrow[rr, "F(\kappa)"] \arrow[dr, "\alpha_i"'] \arrow[ddr, bend right, "\beta_i"'] & & F(j) \arrow[dl, "\alpha_j"] \arrow[ddl, bend left, "\beta_j"] \\ & \operatorname{colim} F \arrow[d, "f"] & \\ & X & \end{tikzcd}$$

for all $\kappa\colon i \to j$ in I.

We have the obvious analogue of Lemma 8.4.5, which again tells us that we may speak of "the" colimit.

LEMMA 8.4.16. *If* $(X,(\beta_i)_{i\in I})$ *and* $(\operatorname{colim} F,(\alpha_i)_{i\in I})$ *are colimits of a functor* $F\colon I\to\mathscr{C}$*, then there is a unique isomorphism* $f\colon \operatorname{colim} F\to X$ *such that* $\alpha_i = f\circ\beta_i$ *for all* $i\in I$*.*

REMARK 8.4.17. When it exists, the colimit of $F\colon I\to\mathscr{C}$ in $\mathscr{C}$ satisfies

$$\operatorname{colim} F = \operatorname{op}(\lim(\operatorname{op}\circ F)),$$

so its underlying object is an limit in $\mathscr{C}^{\mathrm{op}}$.

DEFINITION 8.4.18. The colimit of a functor $F\colon I\to\mathscr{C}$ from a category I with only identity morphisms is called a *coproduct*, and it is denoted $\amalg_{i\in I}F(i)$.

EXAMPLES 8.4.19.

a. The coproduct in **Set** and **Top** of two objects X_1 and X_2 is the disjoint union $X_1\amalg X_2$.

b. The coproduct in **Gp** of two groups G_1 and G_2 is the free product $G_1 * G_2$.

c. The coproduct in R-**mod** (and in particular **Ab**) of two R-modules A_1 and A_2 is the direct sum $A_1\oplus A_2$.

d. The coproduct in the category of commutative rings R_1 and R_2 is the tensor product $R_1\otimes R_2$.

REMARK 8.4.20. Examples 8.4.19(a-d) generalize directly to arbitrary collections of objects.

REMARK 8.4.21. Much as with limits, we may speak of a colimit of a diagram in a category.

DEFINITION 8.4.22. The colimit of a diagram

(8.4.2)
$$\begin{tikzcd}
B \arrow[r, "g_1"] \arrow[d, "g_2"] & A_1 \\
A_2, &
\end{tikzcd}$$

in $\mathscr{C}$ is called the *pushout* $A_1\amalg_B A_2$.

REMARK 8.4.23. The pushout of the diagram (8.4.2) fits into a diagram

$$\begin{tikzcd}
B \arrow[r, "g_1"] \arrow[d, "g_2"] & A_1 \arrow[d, "\iota_1"] \\
A_2 \arrow[r, "\iota_2"] & A_1\amalg_B A_2,
\end{tikzcd}$$

where ι_1 and ι_2 are induced by the universal property of the colimit.

EXAMPLE 8.4.24. In **Set** and **Top**, the pushout is the quotient (set or topological space) of the disjoint union of A_1 and A_2 under the equivalence relation identifying $g_1(b)$ with $g_2(b)$ for all $b\in B$.

DEFINITION 8.4.25. We say that a category $\mathscr{C}$ *admits* the limit (resp., colimit) of a functor $F\colon I \to \mathscr{C}$ if the limit (resp., colimit) exists in $\mathscr{C}$.

REMARK 8.4.26. More generally, we may speak of $\mathscr{C}$ *admitting* the limits (or colimits) of any collection of functors from small categories to $\mathscr{C}$.

DEFINITION 8.4.27. A category is called *complete* if it admits all limits.

EXAMPLE 8.4.28. The category of finite sets is not complete.

DEFINITION 8.4.29. A category is called *cocomplete* if it admits all colimits.

REMARK 8.4.30. To say that $\mathscr{C}$ is complete is to say that $\mathscr{C}^{\mathrm{op}}$ is cocomplete.

PROPOSITION 8.4.31. *The category* **Set** *is both complete and cocomplete.*

PROOF. Let $F\colon I \to$ **Set** be a functor. We merely describe the limit and colimit of F and leave the rest to the reader. The limit is

$$\lim F = \left\{ (a_i)_i \in \prod_{i \in I} F(i) \mid F(\phi)(a_i) = a_j \text{ if } \phi\colon i \to j \text{ in } I \right\},$$

and the colimit is

$$\operatorname{colim} F = \amalg_{i \in I} F(i)/\sim$$

where $\sim$ is the minimal equivalence relation satisfying $a_i \sim a_j$ for $a_i \in F(i)$ and $a_j \in F(j)$ if there exists $\phi\colon i \to j$ with $F(\phi)(a_i) = a_j$. □

REMARK 8.4.32. In fact, the categories **Set**, **Top**, **Gp**, **Ab**, **Ring**, and R-**mod** admit all limits and colimits.

The reader will easily verify the following.

PROPOSITION 8.4.33. *Let I be a small category and $\mathscr{C}$ a (co)complete category. Then the category* **Func**$(I,\mathscr{C})$ *is (co)complete.*

COROLLARY 8.4.34. *Let I be a small category and $\mathscr{C}$ be (co)complete. Let $F\colon I \to \mathscr{C}$ be a functor, and supposing that $\mathscr{C}$ is small, consider the Yoneda embedding $h^{\mathscr{C}}\colon \mathscr{C} \to$* **Func**$(\mathscr{C}^{\mathrm{op}},$ **Set**). *Then $h^{\mathscr{C}} \circ F$ has a (co)limit in* **Func**$(\mathscr{C}^{\mathrm{op}},$ **Set**).

DEFINITION 8.4.35. A *directed set I* is a set I together with a partial ordering $\leq$ on I such that for any $i, j \in I$, there exists $k \in I$ with $i \leq k$ and $j \leq k$.

DEFINITION 8.4.36.

a. The limit of a diagram

$$\cdots \to A_3 \to A_2 \to A_1$$

in a category $\mathscr{C}$ is referred to as the *sequential limit* of the objects A_i.

b. The colimit of a diagram

$$A_1 \to A_2 \to A_3 \to \cdots$$

in a category $\mathscr{C}$ is referred to as the *sequential colimit* of the objects A_i.

EXAMPLE 8.4.37. In **Ab**, the sequential limit of the groups $\mathbb{Z}/p^n\mathbb{Z}$ with respect to homomorphisms $\mathbb{Z}/p^{n+1}\mathbb{Z} \to \mathbb{Z}/p^n\mathbb{Z}$ given by reduction modulo p^n is the group $\mathbb{Z}_p$ of p-adic integers. The sequential colimit of these same groups with respect to the maps $\mathbb{Z}/p^n\mathbb{Z} \to \mathbb{Z}/p^{n+1}\mathbb{Z}$ induced by multiplication modulo p is the group $\mathbb{Q}_p/\mathbb{Z}_p$, equal to the p-power torsion in $\mathbb{Q}/\mathbb{Z}$.

The sequential limit (resp., sequential colimit) is just a special case of the notion of an inverse limit (resp., direct limit), which is a more usual terminology.

DEFINITION 8.4.38.

a. A *directed category* I is a category with a nonempty directed set I of objects and at most one morphism $i \to j$ for any $i, j \in I$, which exists if and only if $i \le j$.

b. A *codirected category* is a category I such that I^{op} is directed.

DEFINITION 8.4.39. Let I be a codirected category. The limit of a functor $F\colon I \to \mathscr{C}$ is referred to the *inverse limit* of the objects $F(i)$ over the inverse system of objects $F(i)$ for $i \in I$ and morphisms $F(\kappa)$ for $\kappa\colon i \to j$ in I, and it is denoted $\varprojlim_{i\in I} F(i)$.

DEFINITION 8.4.40. Let I be a directed category. The colimit of a functor $F\colon I \to \mathscr{C}$ is the *direct limit* of the objects $F(i)$ over the directed system of objects $F(i)$ for $i \in I$ and morphisms $F(\kappa)$ for $\kappa\colon i \to j$ in I and is denoted $\varinjlim_{i\in I} F(i)$ (or sometimes just $\varinjlim F$).

EXAMPLES 8.4.41.

a. The inverse limit of the (commutative) rings $\mathbb{Z}/p^n\mathbb{Z}$ with respect to homomorphisms $\mathbb{Z}/p^{n+1}\mathbb{Z} \to \mathbb{Z}/p^n\mathbb{Z}$ given by reduction modulo p^n is the ring $\mathbb{Z}_p$ known as the p-adic integers.

b. The direct limit of the abelian groups $\mathbb{Z}/p^n\mathbb{Z}$ with respect to the multiplication-by-p maps $\mathbb{Z}/p^n\mathbb{Z} \to \mathbb{Z}/p^{n+1}\mathbb{Z}$ is equal to the subgroup of elements of p-power order (under addition) in $\mathbb{Q}/\mathbb{Z}$.

c. The absolute group $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ of all automorphisms of the field of algebraic numbers $\overline{\mathbb{Q}}$ is isomorphic to the inverse limit of the collection of all $\mathrm{Gal}(K/\mathbb{Q})$ for K a finite, normal extension of $\mathbb{Q}$ inside $\overline{\mathbb{Q}}$, with respect to the maps $\mathrm{Gal}(L/\mathbb{Q}) \to \mathrm{Gal}(K/\mathbb{Q})$ given by restriction with $K \subseteq L$. Note that the set of field extensions is directed as the compositum of two normal extensions is normal, and the resulting category is codirected as we use only the restriction morphisms.

For certain diagrams, limits and colimits can be quite boring, especially when the diagram contains an initial object in the case of limits, or terminal object in the case of colimits.

DEFINITION 8.4.42.

a. An *initial object* A in $\mathscr{C}$ is an object such that for each $B \in \mathrm{Obj}(\mathscr{C})$, there is a unique morphism $A \to B$ in $\mathscr{C}$.

b. A *terminal object* X in $\mathscr{C}$ is an object such that for each $B \in \mathrm{Obj}(\mathscr{C})$, there is a unique morphism $B \to X$ in $\mathscr{C}$.

c. An *zero object* 0 in $\mathscr{C}$ is an object that is both initial and terminal.

REMARK 8.4.43. Terminal and initial objects are unique up to unique isomorphism when defined.

We provide some examples.

EXAMPLES 8.4.44.

a. The empty set $\varnothing$ is the initial object in the category **Set**, while any set with one element is a terminal object.

b. The trivial group is a zero object in the category **Gp**.

c. The zero ring is a terminal object and $\mathbb{Z}$ is an initial object in **Ring**.

We omit the proof of the following easy lemma.

LEMMA 8.4.45. *Let I be a small category and $F\colon I \to \mathscr{C}$ a functor.*

a. Suppose that I has an initial object i. Then

$$\lim F = F(i).$$

b. Suppose that I has a terminal object j. Then

$$\operatorname{colim} F = F(j).$$

8.5. Adjoint functors

The following definition allows us to weaken the property of being quasi-inverse to one of "adjointness".

DEFINITION 8.5.1. We say that $F\colon \mathscr{C} \to \mathscr{D}$ is *left adjoint* to $G\colon \mathscr{D} \to \mathscr{C}$ if there exist bijections

$$\phi_{C,D}\colon \operatorname{Hom}_{\mathscr{D}}(F(C),D) \xrightarrow{\sim} \operatorname{Hom}_{\mathscr{C}}(C,G(D))$$

for each $C \in \operatorname{Obj}(\mathscr{C})$ and $D \in \operatorname{Obj}(\mathscr{D})$ such that the $\phi_{C,D}$ form a natural transformation of functors

$$\mathscr{C}^{\mathrm{op}} \times \mathscr{D} \to \mathbf{Set}.$$

We also say that G is *right adjoint* to F, and we say that F and G are *adjoint functors*.

REMARK 8.5.2. To say that η is a natural transformation in Definition 8.5.1 is a fancier way of saying that given morphisms $f\colon C' \to C$ in $\mathscr{C}$ and $g\colon D \to D'$ in $\mathscr{D}$, we have a commutative diagram

$$\begin{tikzcd}
\operatorname{Hom}_{\mathscr{D}}(F(C),D) \arrow[r, "\eta_{(C,D)}"] \arrow[d, "t \mapsto g \circ t \circ F(f)"] & \operatorname{Hom}_{\mathscr{C}}(C,G(D)) \arrow[d, "u \mapsto G(g) \circ u \circ f"] \\
\operatorname{Hom}_{\mathscr{D}}(F(C'),D') \arrow[r, "\eta_{(C',D')}"] & \operatorname{Hom}_{\mathscr{C}}(C',G(D')).
\end{tikzcd}$$

REMARK 8.5.3. If $F\colon \mathscr{C} \to \mathscr{G}$ and $G\colon \mathscr{D} \to \mathscr{C}$ are quasi-inverse functors, then we have bijections

$$\phi_{C,D}\colon \operatorname{Hom}_{\mathscr{D}}(F(C),D) \xrightarrow{\sim} \operatorname{Hom}_{\mathscr{C}}(C,G(D)), \qquad \phi_{C,D}(g) = G(g) \circ \eta_C,$$

for $C \in \operatorname{Obj}(\mathscr{C})$ and $D \in \operatorname{Obj}(\mathscr{D})$, where $g\colon F(C) \to D$ in $\mathscr{D}$ and η is a natural isomorphism $G \circ F \rightsquigarrow \mathrm{id}_{\mathscr{C}}$. These form a natural transformation ϕ between functors $\mathscr{C}^{\mathrm{op}} \times \mathscr{D} \to \mathbf{Set}$, so G is right adjoint to F. Similarly, using $\eta'\colon F \circ G \rightsquigarrow \mathrm{id}_{\mathscr{D}}$, we see that G is left adjoint to F.

EXAMPLE 8.5.4. The forgetful functor $\mathbf{Gp} \to \mathbf{Set}$ is right adjoint to the functor $\mathbf{Set} \to \mathbf{Gp}$ that takes a set X to the free group F_X on X and a map $f\colon X \to Y$ of sets to the unique homomorphism $\phi_f\colon F_X \to F_Y$ with $\phi_f|_X = f$. That is, the restriction map

$$\mathrm{Hom}(F_Y, G) \to \mathrm{Maps}(Y, G)$$

is inverse to the map $\mathrm{Maps}(Y,G) \to \mathrm{Hom}(F_Y,G)$ that takes $\phi_f\colon Y \to G$ to $F(f)$, and these bijections are easily seen to be natural.

Later, we will treat what is perhaps the most standard example of adjointness: that of Hom and $\otimes$ in categories of modules.

PROPOSITION 8.5.5. *Fix categories I and $\mathscr{C}$, and suppose that all limits $F\colon I \to \mathscr{C}$ exist. The functor* lim *has a left adjoint Δ given by taking $A \in \mathrm{Obj}(\mathscr{C})$ to the constant functor c_A, where $c_A(i) = A$ for all $i \in I$, and taking $g\colon A \to B$ for $A, B \in \mathrm{Obj}(\mathscr{C})$ to the natural transformation $c_A \rightsquigarrow c_B$ given by $g\colon c_A(i) = A \to c_B(i) = B$ for all $i \in I$.*

PROOF. We must describe natural isomorphisms

$$\mathrm{Hom}_{\mathbf{Func}(I,\mathscr{C})}(c_A, F) \cong \mathrm{Hom}_{\mathscr{C}}(A, \lim F)$$

for $A \in \mathrm{Obj}(\mathscr{C})$ and $F\colon I \to \mathscr{C}$. I.e., given a natural tranformation $\eta\colon c_A \rightsquigarrow F$, we must associate a map $f\colon A \to \lim F$, and conversely. Such a natural transformation η consists of maps

$$\eta_i\colon c_A(i) = A \to F(i)$$

that are compatible in the sense that $\eta_j = F(\kappa) \circ \eta_i$ for all $\kappa\colon i \to j$. Thus, the existence of a unique f is simply the universal property of the limit. On the other hand, if we have f, then we have maps

$$\phi_i \circ f\colon A \to F(i),$$

where ϕ_i is the map $\lim F \to F(i)$ arising in the definition of the limit. These maps then define the universal transformation η. □

We now see exactly how adjointness weakens inverseness.

DEFINITION 8.5.6. Two categories $\mathscr{C}$ and $\mathscr{D}$ are said to be *equivalent* if there exist functors $F\colon \mathscr{C} \to \mathscr{D}$ and $G\colon \mathscr{D} \to \mathscr{C}$ and natural isomorphisms $\eta\colon G \circ F \rightsquigarrow \mathrm{id}_{\mathscr{C}}$ and $\eta'\colon F \circ G \rightsquigarrow \mathrm{id}_{\mathscr{D}}$. Two such functors F and G are said to be *quasi-inverse*, and F and G are said to be *equivalences of categories*.

EXAMPLE 8.5.7. A category $\mathscr{C}$ with one object 0 and one morphism is equivalent to the category $\mathscr{D}$ with two objects 1, 2 and four morphisms, the identity morphisms of 1 and 2 and isomorphisms $1 \to 2$ and $2 \to 1$. We have quasi-inverse functors F and G with $F(0) = 1$ and $F(\mathrm{id}_0) = \mathrm{id}_1$ and $G(1) = G(2) = 0$ and $G(f) = \mathrm{id}_0$ for all f. To see naturality, note that every morphism between two objects in either category is unique.

NOTATION 8.5.8. Let $\eta\colon F \rightsquigarrow F'$ be a natural transformation between functors $F, F'\colon \mathscr{C} \to \mathscr{D}$.

a. If $G\colon \mathscr{D} \to \mathscr{E}$ is a functor, then we define a natural transformation $G(\eta)\colon G \circ F \rightsquigarrow G \circ F'$ by

$$G(\eta)_C = G(\eta_C)\colon G(F(C)) \to G(F'(C))$$

for all objects C of $\mathscr{C}$.

b. If $H\colon \mathscr{B} \to \mathscr{C}$ is a functor, then we define a natural transformation $\eta(H)\colon F \circ H \rightsquigarrow F' \circ H$ by

$$\eta(H)_B = \eta_{H(B)}\colon F(H(B)) \to F'(H(B))$$

for all objects B of $\mathscr{C}$.

DEFINITION 8.5.9. Let $F\colon \mathscr{C} \to \mathscr{D}$ and $G\colon \mathscr{D} \to \mathscr{C}$ be functors.

a. A *unit* for the pair (F,G) is a natural transformation $\mathrm{id}_{\mathscr{C}} \rightsquigarrow G \circ F$.

b. A *counit* for the pair (F,G) is a natural transformation $F \circ G \rightsquigarrow \mathrm{id}_{\mathscr{D}}$.

c. A *unit-counit adjunction* is a pair (F,G), a unit η for (F,G), and a counit η' for (F,G) satisfying

$$\mathrm{id}_F = \eta'(F) \circ F(\eta)\colon F \rightsquigarrow F$$

as morphisms in $\mathbf{Func}(\mathscr{C},\mathscr{D})$ and

$$\mathrm{id}_G = G(\eta') \circ \eta(G)\colon G \rightsquigarrow G$$

as morphisms in $\mathbf{Func}(\mathscr{D},\mathscr{C})$.

PROPOSITION 8.5.10. *A functor $F\colon \mathscr{C} \to \mathscr{D}$ is left adjoint to a functor $G\colon \mathscr{D} \to \mathscr{C}$ if and only if there exists a unit-counit adjunction for the pair (F,G).*

PROOF. Suppose that F is left adjoint to G. We define $\eta\colon \mathrm{id}_{\mathscr{C}} \rightsquigarrow G \circ F$ as follows. For $C \in \mathrm{Obj}(\mathscr{C})$, we have bijections

$$\mathrm{Hom}_{\mathscr{D}}(F(C),F(C)) \xrightarrow{\sim} \mathrm{Hom}_{\mathscr{C}}(C, G \circ F(C))$$

by adjointness, and we define η_C to be the image of $\mathrm{id}_{F(C)}$. For $D \in \mathrm{Obj}(\mathscr{D})$, we also have

$$\mathrm{Hom}_{\mathscr{D}}(F \circ G(D), D) \xrightarrow{\sim} \mathrm{Hom}_{\mathscr{C}}(G(D), G(D))$$

and define $\eta'\colon F \circ G \rightsquigarrow \mathrm{id}_{\mathscr{D}}$ by taking η'_D to be the image of $\mathrm{id}_{G(D)}$ under the inverse of this map. We leave it to the reader to check that these are natural and form a unit-counit adjunction. The converse is left to the reader as well. □

8.6. Representable functors

DEFINITION 8.6.1. Let $F\colon \mathscr{C} \to \mathbf{Set}$ be a contravariant functor. Then F is said to be *representable* if there exists a natural isomorphism $h^B \rightsquigarrow F$ for some $B \in \mathrm{Obj}(\mathscr{C})$. (In other words, we have natural bijections

$$\mathrm{Hom}_{\mathscr{C}}(A,B) \xrightarrow{\sim} F(A)$$

for all objects A of $\mathscr{C}$.) We then say that B *represents* F.

Using Yoneda's lemma and assuming $\mathscr{C}$ to be small, we can reword Definition 8.6.1 as saying that there exists $B \in \mathrm{Obj}(\mathscr{C})$ such that there are compatible bijections between the set of natural transformations $h^A \rightsquigarrow F$ and the set of morphisms $A \to B$ for each $A \in \mathrm{Obj}(\mathscr{C})$.

EXAMPLE 8.6.2. Consider the contravariant functor $P\colon \mathbf{Set} \to \mathbf{Set}$ which takes a set S to its *power set* $P(S)$, the set of all subsets of S and a map $f\colon S \to T$ to the map $P(f)\colon P(T) \to P(S)$ by mapping $U \subset T$ to $f^{-1}(U)$. Then P is represented by the set $\{0,1\}$ via the isomorphism

$$\mathrm{Maps}(S,\{0,1\}) \xrightarrow{\sim} P(S)$$

by $\phi \mapsto \phi^{-1}(\{1\})$. These isomorphisms form a natural transformation:

$$\begin{tikzcd}
\mathrm{Maps}(T,\{0,1\}) \arrow[r, "\sim"] \arrow[d, "h^{\{0,1\}}(f)"] & P(T) \arrow[d, "P(f)"] \\
\mathrm{Maps}(S,\{0,1\}) \arrow[r, "\sim"] & P(S)
\end{tikzcd}$$

for $f\colon S \to T$. Here, the lefthand vertical map takes ϕ to $\phi \circ f$ and the righthand vertical map takes a subset X of T to $f^{-1}(X)$. We check that

$$(\phi \circ f)^{-1}(\{1\}) = f^{-1}(\phi^{-1}(\{1\})).$$

The following is a corollary of Yoneda's lemma.

LEMMA 8.6.3. *A representable functor is represented by a unique object up to isomorphism. If B and C represent a contravariant functor $F\colon \mathscr{C} \to$ **Set**, then such an isomorphism $f\colon B \to C$ is unique making the diagrams*

$$\begin{tikzcd}
\mathrm{Hom}_{\mathscr{C}}(A,B) \arrow[r, "\sim"] \arrow[d, "h^A(f)"] & F(A) \arrow[d, equal] \\
\mathrm{Hom}_{\mathscr{C}}(A,C) \arrow[r, "\sim"] & F(A)
\end{tikzcd}$$

commute for all $A \in \mathrm{Obj}(\mathscr{C})$.

PROOF. Let $F\colon \mathscr{C} \to \mathbf{Set}$ be a representable (contravariant) functor represented by $B \in \mathrm{Obj}(\mathscr{C})$ and $C \in \mathrm{Obj}(\mathscr{C})$. Then we have natural isomorphisms $\xi\colon h^B \rightsquigarrow F$ and $\xi'\colon h^C \rightsquigarrow F$. The composition $\xi' \circ \xi^{-1}\colon h^B \to h^C$ is equal to $h^{\mathscr{C}}(f)$ for a unique $f\colon B \to C$ by the weak form of Yoneda's lemma. □

THEOREM 8.6.4. *Let $F\colon \mathscr{C} \to \mathscr{D}$ be a functor between small categories.*

a. The functor F has a right adjoint if and only if the functor $h^D \circ F$ is representable for each $D \in \mathrm{Obj}(\mathscr{D})$. If G is right adjoint to F, then $h^D \circ F$ is representable by $G(D)$.

b. If F has a right adjoints G and G', then there exists a unique natural isomorphism $\xi\colon G \rightsquigarrow G'$ such that diagrams

$$\begin{tikzcd}
\mathrm{Hom}_{\mathscr{D}}(F(C),D) \arrow[r, "\eta_{(C,D)}"] \arrow[d, equal] & \mathrm{Hom}_{\mathscr{C}}(C,G(D)) \arrow[d, "f\mapsto \xi_D\circ f"] \\
\mathrm{Hom}_{\mathscr{D}}(F(C),D) \arrow[r, "\eta'_{(C,D)}"] & \mathrm{Hom}_{\mathscr{C}}(C,G'(D))
\end{tikzcd}$$

commute for all $C \in \mathrm{Obj}(\mathscr{C})$ and $D \in \mathrm{Obj}(\mathscr{D})$, where the horizontal morphisms are the adjunction isomorphisms.

PROOF. Assume that F has a right adjoint G, and consider the adjunction morphisms

$$\eta_{(C,D)}\colon\ \mathrm{Hom}_{\mathscr{D}}(F(C),D) \xrightarrow{\sim} \mathrm{Hom}_{\mathscr{C}}(C,G(D)).$$

In other words,

$$h^D \circ F(C) \cong h^{G(D)}(C),$$

so $G(D)$ represents $h^D \circ F$. In this case, the uniqueness in part b is an immediate consequence of Lemma 8.6.3.

Now suppose that $h^D \circ F$ is representable for each D by some object $G(D)$ (chosen using the axiom of choice). Then there exist isomorphisms $\eta_{(C,D)}$ that are natural in C. We must also define G on morphisms $f\colon D \to D'$ in $\mathscr{D}$. Such an f induces a natural transformation $h^D \rightsquigarrow h^{D'}$ which provides morphisms

$$h^D \circ F(C) \to h^{D'} \circ F(C)$$

for all $C \in \mathrm{Obj}(\mathscr{C})$ and thus induces $h^{G(D)}(C) \to h^{G(D')}(C)$, and these are natural in C. Thus, we have a natural transformation $h^{G(D)} \rightsquigarrow h^{G(D')}$. Since the Yoneda embedding is fully faithful, we have a unique morphism $G(D) \to G(D')$ inducing this natural transformation, which we define to be $G(f)$. We leave to the reader the check that G as defined is a functor. □

DEFINITION 8.6.5. Let $F\colon \mathscr{C} \to \mathbf{Set}$ be a covariant functor. We say that F is *representable* if there exists a natural isomorphism $h_A \rightsquigarrow F$ for some $A \in \mathrm{Obj}(\mathscr{C})$. (That is, there are natural isomorphisms

$$F(B) \xrightarrow{\sim} \mathrm{Hom}_{\mathscr{C}}(A,B)$$

in $B \in \mathrm{Obj}(\mathscr{C})$.) In this case, we say that A *represents* F.

REMARK 8.6.6. A covariant functor $F\colon \mathscr{C} \to \mathbf{Set}$ is representable if and only if the contravariant functor $F \circ \mathrm{op}\colon \mathscr{C}^{\mathrm{op}} \to \mathbf{Set}$ is representable. The same object of $\mathscr{C}$ will represent both objects.

EXAMPLE 8.6.7. Let $F\colon \mathbf{Gp} \to \mathbf{Set}$ be the forgetful functor. Then F can be represented by $\mathbb{Z}$. To see this, we define the set map

$$G \to \mathrm{Hom}_{\mathbf{Gp}}(\mathbb{Z},G)$$

by $a \mapsto (1 \mapsto a)$ for $a \in G$. Naturality is clear.

EXAMPLE 8.6.8. Let $F\colon \mathbf{Gp} \to \mathbf{Set}$ be the functor which sends a group to its subset $G[n]$ of elements of order dividing n. Then F can be represented by $\mathbb{Z}/n\mathbb{Z}$.

EXAMPLE 8.6.9. Consider a functor $F\colon I \to \mathscr{C}$ between small categories. To say that the contravariant functor $\lim h^{\mathscr{C}} \circ F\colon \mathscr{C} \to \mathbf{Set}$ is representable is exactly to say that there exists an object X in $\mathscr{C}$ such that one has natural isomorphisms

$$\mathrm{Hom}_{\mathscr{C}}(A,X) \xrightarrow{\sim} \lim(h^{\mathscr{C}} \circ F)(A) \xrightarrow{\sim} \lim(\mathrm{Hom}_{\mathscr{C}}(A,F(\cdot)))$$

for $A \in \mathscr{C}$. In other words, $\lim h^{\mathscr{C}} \circ F$ is representable if and only if $\lim F$ exists in $\mathscr{C}$.

EXAMPLE 8.6.10. Consider a functor $F\colon I \to \mathscr{C}$ between small categories. View $h_{\mathscr{C}}$ as a covariant functor $\mathscr{C} \to \mathrm{Hom}(\mathscr{C},\mathbf{Set})$. To say that the functor $\lim(h_{\mathscr{C}} \circ F)\colon \mathscr{C} \to \mathbf{Set}$ is representable is exactly to say that there exists an object $X \in \mathscr{C}$ such that one has natural isomorphisms

$$\mathrm{Hom}_{\mathscr{C}}(X,A) \xrightarrow{\sim} \lim(h_{\mathscr{C}} \circ F)(A) \xrightarrow{\sim} \lim(\mathrm{Hom}_{\mathscr{C}}(F(\cdot),A))$$

for $A \in \mathscr{C}$. In other words, $\lim h_{\mathscr{C}} \circ F$ is representable if and only if $\operatorname{colim} F$ exists in $\mathscr{C}$.

8.7. Equalizers and images

DEFINITION 8.7.1. Let $\mathscr{C}$ be a category, and let

$$\begin{tikzcd} A \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & B. \end{tikzcd} \tag{8.7.1}$$

be a diagram in $\mathscr{C}$.

a. The limit $\mathrm{eq}(f,g)$ of the diagram (8.7.1), when it exists, is called its *equalizer*.

b. The colimit $\mathrm{coeq}(f,g)$ of (8.7.1) is called its *coequalizer*.

We have a commutative diagram:

$$\begin{tikzcd} \mathrm{eq}(f,g) \arrow[r] & A \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & B \arrow[r] & \mathrm{coeq}(f,g). \end{tikzcd}$$

EXAMPLES 8.7.2.

a. Let X,Y be sets, and consider maps $f,g\colon X \to Y$. In $\mathbf{Set}$, we have

$$\mathrm{eq}(f,g) = \{x \in X \mid f(x) = g(x)\}$$

and $\mathrm{coeq}(f,g)$ is the quotient of Y by the minimal equivalence relation $\sim$ generated by $f(x) \sim g(x)$ for all $x \in X$.

b. In R-**mod**, the equalizer is expressed as in $\mathbf{Set}$. For an R-module homomorphism $f\colon A \to B$, we have

$$\mathrm{coeq}(f,g) = B/\{(f-g)(a) \mid a \in A\}.$$

LEMMA 8.7.3. *Let $f,g\colon A \to B$ be morphisms in a category $\mathscr{C}$.*

a. Suppose that $\mathrm{eq}(f,g)$ *exists. Then the induced map $h\colon$* $\mathrm{eq}(f,g) \to A$ *is a monomorphism.*

b. Suppose that $\mathrm{coeq}(f,g)$ *exists. Then the induced map $k\colon B \to$* $\mathrm{coeq}(f,g)$ *is an epimorphism.*

PROOF. Suppose that $\alpha, \beta \colon C \to \mathrm{eq}(f,g)$ are morphisms in $\mathscr{C}$ such that $h \circ \alpha = h \circ \beta$. Let $h' = h \circ \alpha$, and note that $f \circ h' = g \circ h'$. But then $\alpha \colon C \to \mathrm{eq}(f,g)$ is unique such that $h' = h \circ \alpha$ by the universal property of $\mathrm{eq}(f,g)$. Since $h' = h \circ \beta$ as well, we have $\alpha = \beta$. Part b follows from part a by working in the opposite category. □

THEOREM 8.7.4. *A category that admits all products and equalizers is complete, and a category that admits all coproducts and coequalizers is cocomplete.*

PROOF. For the second statement, by taking the opposite category, we are reduced to the first statement. Let $\mathscr{C}$ be a category that admits all products and equalizers. Let $F \colon I \to \mathscr{C}$ be a functor from a small category I, and consider the equalizer $\mathrm{eq}(f,g)$ of the two morphisms

$$f, g \colon \prod_{i \in I} F(i) \to \prod_{\phi \colon i \to \phi(i)} F(\phi(i))$$

defined via the universal property of the second product as the unique morphisms satisfying

$$p_\phi \circ f = p_{\phi(i)} \colon \prod_{j \in I} F(j) \to F(\phi(i))$$

and

$$p_\phi \circ g = F(\phi) \circ p_i \colon \prod_{j \in I} F(j) \to F(\phi(i))$$

for morphisms $\phi \colon i \to \phi(i)$ in $\mathscr{C}$, where p_ϕ denotes projection onto the ϕ-coordinate in the second product and p_i denotes projection to the i-coordinate in the first.

Let $\iota \colon \mathrm{eq}(f,g) \to \prod_{i \in I} F(i)$ be the morphism defining the equalizer. We claim that the equalizer $\mathrm{eq}(f,g)$, together with the maps $p_i \circ \iota \colon \mathrm{eq}(f,g) \to F(i)$ for $i \in I$, satisfies the univeral property of $\lim F$. By definition, for any morphism ϕ in $\mathscr{C}$ as above, we have

$$F(\phi) \circ (p_i \circ \iota) = p_\phi \circ g \circ \iota = p_\phi \circ f \circ \iota = p_{\phi(i)} \circ \iota.$$

Moreover, given $X \in \mathscr{C}$ and morphisms $\psi_i \colon X \to F(i)$ for $i \in I$ such that $F(\phi) \circ \psi_i = \phi_{\phi(i)}$ for all $i \in I$, we have a product map $\psi \colon X \to \prod_{i \in I} F(i)$ such that

$$p_\phi \circ f \circ \psi = p_{\phi(i)} \circ \psi = \psi_{\phi(i)}$$

and

$$p_\phi \circ g \circ \psi = F(\phi) \circ p_i \circ \psi = F(\phi) \circ \psi_i = \psi_{\phi(i)},$$

so there exists a unique morphism $\theta \colon X \to \mathrm{eq}(f,g)$ with $p_i \circ \iota \circ \theta = \psi_i$ for all $i \in I$. That is, $\mathrm{eq}(f,g)$ satisfies the universal property of the limit. □

DEFINITION 8.7.5. In a category $\mathscr{C}$ with a zero object 0, the *zero morphism* $0 \colon A \to B$ between objects $A, B \in \mathrm{Obj}(\mathscr{C})$ is the composition of $A \to 0 \to B$ of morphisms proscribed by the fact that 0 is both initial and terminal.

DEFINITION 8.7.6. Let $\mathscr{C}$ be a category with a zero object. Let $f \colon A \to B$ be a morphism in $\mathscr{C}$, and let $0 \colon A \to B$ be the zero morphism.

a. The *kernel* $\ker f$ of f is the equalizer of f and 0.

b. The *cokernel* $\mathrm{coker} f$ of f is the coequalizer of f and 0.

EXAMPLES 8.7.7.

a. In R-**mod**, the categorical notions of kernel and cokernel agree with the usual ones.

b. In **Gp**, kernel is the usual notion, and the cokernel of a group homomorphism $f\colon G \to H$ with the quotient of H by the normal closure of $f(G)$.

There are different notions of image and coimages in categories. We use the following.

DEFINITION 8.7.8. Let $f\colon A \to B$ be a morphism in a category $\mathscr{C}$.

a. The *image* of f is an object $\operatorname{im} f$ of $\mathscr{C}$ together with a monomorphism $\iota\colon \operatorname{im} f \to B$ such that there exists a morphism $\pi\colon A \to \operatorname{im} f$ such that $\pi \circ \iota = f$ and such that if $g\colon C \to B$ us a monomorphism and $e\colon A \to C$ is a morphism such that $e \circ g = f$, then there exists a unique morphism $\psi\colon \operatorname{im} f \to C$ such that $g = \iota \circ \psi$.

b. The *coimage* of f is an object $\operatorname{coim} f$ together with an epimorphism $\pi\colon A \to \operatorname{coim} f$ such that there exists a morphism $\iota\colon \operatorname{coim} f \to B$ with $\iota \circ \pi = f$ and such that if $e\colon A \to C$ is an epimorphism and $g\colon C \to B$ is a morphism such that $f = e \circ g$, the there exists a unique morphism $\theta\colon \operatorname{coim} f \to C$ such that $e = \theta \circ f$.

REMARK 8.7.9. In the definition of the image of $f\colon A \to B$, then π is uniquely determined as ι is a monomorphism and $f = \iota \circ \pi$. Moreover, if $g \circ \psi = \iota$ as stated, then $g \circ \psi \circ \pi = \iota \circ \pi = g \circ e$, and g is a monomorphism, so $\psi \circ \pi = e$. That is the diagram

$$\begin{tikzcd}
A \arrow[rr, "f"] \arrow[dr, "\pi"] \arrow[ddr, bend right, "e"'] & & B \\
& \operatorname{im} f \arrow[ur, "\iota"'] \arrow[d, "\psi"] & \\
& C \arrow[uur, bend right, "g"'] &
\end{tikzcd}$$

commutes. Finally, note that ψ is forced to be a monomorphism since ι is. The analogous statements hold for the coimage.

PROPOSITION 8.7.10. *Let $f\colon A \to B$ be a morphism in a category $\mathscr{C}$.*

a. If $\mathscr{C}$ admits equalizers, then the canonical morphism $\pi\colon A \to \operatorname{im} f$ is an epimorphism.

b. If $\mathscr{C}$ admits coequalizers, then the canonical morphism $\iota\colon \operatorname{coim} f \to B$ is a monomorphism.

PROOF. We prove only part a, as part b is proven similarly. Suppose that $\alpha, \beta\colon \operatorname{im} f \to D$ satisfy $\alpha \circ \pi = \beta \circ \pi$. Then by definition of the equalizer, there exists a unique morphism $\rho\colon A \to \operatorname{eq}(\alpha, \beta)$ such that for the canonical monomorphism $c\colon \operatorname{eq}(\alpha, \beta) \to \operatorname{im} f$, we have $\pi = c \circ \rho$. On the other hand, consider the composite monomorphism $g = \iota \circ c\colon \operatorname{eq}(\alpha, \beta) \to B$, where $\iota\colon \operatorname{im} f \to B$ is the morphism of the image. Note that $g \circ \rho = f$, so by definition of the image, there exists a unique morphism $d\colon \operatorname{im} f \to \operatorname{eq}(\alpha, \beta)$ such that $g \circ d = \iota$. Then $\iota \circ (c \circ d) = \iota$, so $c \circ d = \mathrm{id}$. On the other hand,

$$g \circ (d \circ c) = \iota \circ c \circ d \circ c = \iota \circ c = g.$$

As g is a monomorphism, we have $d \circ c = \mathrm{id}$ as well. □

REMARK 8.7.11. If $\mathscr{C}$ has finite products, finite coproducts, equalizers, and coequalizers, we may also make the following definition of the image an coimage of a morphism $f\colon A \to B$.

a. The *image* of f is the equalizer of the two morphisms $\iota_i\colon B \to B \amalg_A B$.

b. The *coimage* of f is the coequalizer of the two projection morphisms $p_i\colon A \times_B A \to A$.

This definition agrees with that already given if every morphism in $\mathscr{C}$ factors through an equalizer morphism and $\mathscr{C}$ admits finite limits and colimits. We omit the nontrivial proof.

LEMMA 8.7.12. *For any $f\colon A \to B$ in a category $\mathscr{C}$ that admits equalizers (or coequalizers) and for which* $\operatorname{coim} f$ *and* $\operatorname{im} f$ *exist, there is a unique morphism $u\colon \operatorname{coim} f \to \operatorname{im} f$ such that the composition*

$$A \xrightarrow{s} \operatorname{coim} f \xrightarrow{u} \operatorname{im} f \xrightarrow{t} B$$

of induced morphisms is f.

PROOF. We suppose that $\mathscr{C}$ admits equalizers. By Proposition 8.7.10, the canonical morphism $\pi\colon A \to \operatorname{im} f$ with $t \circ \pi = f$ is an epimorphism. By the definition of $\operatorname{coim} f$, there then exists a unique morphism $u\colon \operatorname{coim} f \to \operatorname{im} f$ such that $\pi = u \circ s$. Then $t \circ u \circ s = t \circ \pi = f$. If $v\colon \operatorname{coim} f \to \operatorname{im} f$ also satisfies $t \circ v \circ s = f$, then $t \circ v \circ s = t \circ \pi$, and $t\colon \operatorname{im} f \to B$ is a monomorphism, so $v \circ s = \pi$. Thus $u = v$ by the uniqueness of u. □

DEFINITION 8.7.13. We say that a morphism $f\colon A \to B$ in a category that admits an image and coimage of f is *strict* if the induced morphism $\operatorname{coim} f \to \operatorname{im} f$ is an isomorphism.

EXAMPLE 8.7.14. Every morphism in the category of R-modules is strict.

8.8. Additive and abelian categories

DEFINITION 8.8.1. An *additive category* $\mathscr{C}$ is a category with the following properties:

i. for $A, B \in \mathrm{Obj}(\mathscr{C})$, the set of morphisms $\mathrm{Hom}_{\mathscr{C}}(A,B)$ in $\mathscr{C}$ has an abelian group law (addition) with the property that for any diagram

$$\begin{tikzcd} A \arrow[r, "f"] & B \arrow[r, shift left, "g_1"] \arrow[r, shift right, "g_2"'] & C \arrow[r, "h"] & D, \end{tikzcd}$$

in $\mathscr{C}$, we have

$$h \circ (g_1 + g_2) \circ f = h \circ g_1 \circ f + h \circ g_2 \circ f,$$

ii. $\mathscr{C}$ has a zero object 0,

iii. $\mathscr{C}$ admits finite coproducts.

In an additive category $\mathscr{C}$, there always exists the zero morphism is the identity element in the abelian group $\mathrm{Hom}_{\mathscr{C}}(A,B)$.

EXAMPLES 8.8.2.

a. The categories **Ab** and R-**mod** are additive categories, with the usual addition of homomorphisms.

b. The full subcategory R-**mod** of finitely generated R-modules is an additive category.

In an additive category, we denote the coproduct of two objects A_1 and A_2 by $A_1 \oplus A_2$.

LEMMA 8.8.3. *Finite products exist in an additive category, and there are natural isomorphisms*

$$A_1 \amalg A_2 \cong A_1 \times A_2$$

for $A_1, A_2 \in \mathrm{Obj}(\mathscr{C})$. *The resulting inclusion morphisms* $\iota_i \colon A_i \to A_1 \amalg A_2$ *and projection morphisms and* $p_i \colon A_1 \amalg A_2 \to A_i$ *obtained by viewing* $A_1 \amalg A_2$ *as a product and coproduct, respectively, satisfy* $p_i \circ \iota_i = \mathrm{id}_{A_i}$ *and* $p_i \circ \iota_j = 0$ *for* $i \neq j$, *while*

$$\iota_1 \circ p_1 + \iota_2 \circ p_2 = \mathrm{id}_{A_1 \amalg A_2}.$$

PROOF. We have morphisms $\iota_i \colon A_i \to A_1 \amalg A_2$ by definition. We also have maps $p_i \colon A_1 \amalg A_2 \to A_i$ defined by

$$p_i \circ \iota_j = \begin{cases} \mathrm{id}_{A_i} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

and the universal property of the coproduct. We then have

$$(\iota_1 \circ p_1 + \iota_2 \circ p_2) \circ \iota_i = \iota_i,$$

and hence $\iota_1 \circ p_1 + \iota_2 \circ p_2 = \mathrm{id}_{A_1 \amalg A_2}$, again by the universal property.

Given an object $B \in \mathrm{Obj}(\mathscr{C})$ and morphisms $g_i \colon B \to A_i$, we then have a morphism

$$\psi = \iota_1 \circ g_1 + \iota_2 \circ g_2 \colon B \to A_1 \amalg A_2,$$

which is unique such that

$$p_i \circ \psi = g_i.$$

Hence $A_1 \amalg A_2$ satisfies the universal property of the product. □

DEFINITION 8.8.4. An object A in an additive category $\mathscr{C}$ together with objects A_i, inclusion morphisms $\iota_i \colon A_i \to A$, and projection morphisms $p_i \colon A \to A_i$ for $i \in \{1,2\}$ for which $p_i \circ \iota_j$ is zero if $i \neq j$ and id_{A_i} if $i = j$ and for which $\iota_1 \circ p_1 + \iota_2 \circ p_2 = \mathrm{id}_A$ is called a *biproduct* of the objects A_1 and A_2, and we write it as $A_1 \oplus A_2$.

The notion of a biproduct allows us to reinterpret addition in an additive category. First, note the following definitions.

DEFINITION 8.8.5. Let A be an object in an additive category $\mathscr{C}$.

a. The *diagonal morphism* $\Delta_A \colon A \to A \oplus A$ in $\mathscr{C}$ is the unique morphism induced by two copies of $\mathrm{id}_A \colon A \to A$ and the universal property of the product.

b. The *codiagonal morphism* $\nabla_A \colon A \oplus A \to A$ in $\mathscr{C}$ is the unique morphisms induced by two copies of $\mathrm{id}_A \colon A \to A$ and the universal property of the coproduct.

DEFINITION 8.8.6. Let $\mathscr{C}$ be an additive category, and let $f_1 \colon A_1 \to B_1$ and $f_2 \colon A_2 \to B_2$ be morphisms in $\mathscr{C}$. The *biproduct*, or *direct sum*, $f_1 \oplus f_2$ of the maps f_1 and f_2 is the morphism $A_1 \oplus A_2 \to B_1 \oplus B_2$ induced as the (morphism defined by the universal property of the) coproduct of the composite maps $A_i \to B_i \to B_1 \oplus B_2$, the latter morphisms being inclusions.

REMARK 8.8.7. Equivalently, the direct sum of f_1 and f_2 as in Definition 8.8.6 is induced as the product of the composite maps $A_1 \oplus A_2 \to A_i \to B_i$, the initial morphisms being projections.

Of course, we could make these definitions in an arbitrary category using products and coproducts.

LEMMA 8.8.8. *Let $f, g\colon A \to B$ be two morphisms in an additive category $\mathscr{C}$. Then we have*

$$f + g = \nabla_B \circ (f \oplus g) \circ \Delta_A.$$

PROOF. Let ι_i^A and p_i^A respectively denote the inclusion maps and projection maps for the biproduct $A \oplus A$, and similarly for B. We have

$$\begin{aligned}\nabla_B \circ (f \oplus g) \circ \Delta_A &= \nabla \circ (f \oplus g) \circ (\iota_1^A \circ p_1^A + \iota_2^A \circ p_2^A) \circ \Delta_A \\ &= \nabla_B \circ (f \oplus g) \circ \iota_1^A \circ p_1^A \circ \Delta_A + \nabla_B \circ (f \oplus g) \circ \iota_2^A \circ p_2^A \circ \Delta_A.\end{aligned}$$

Taking the first term without loss of generality, we have

$$\nabla_B \circ ((f \oplus g) \circ \iota_1^A) \circ (p_1^A \circ \Delta_A) = \nabla_B \circ (\iota_1^B \circ f) \circ \mathrm{id}_A = \mathrm{id}_B \circ f = f.$$

□

DEFINITION 8.8.9. A functor $F\colon \mathscr{C} \to \mathscr{D}$ between additive categories is called *additive* if for each $A, B \in \mathrm{Obj}(\mathscr{C})$, the map

$$\mathrm{Hom}_{\mathscr{C}}(A, B) \to \mathrm{Hom}_{\mathscr{D}}(F(A), F(B))$$

is a group homomorphism.

EXAMPLE 8.8.10. Let $\mathscr{C}$ be an additive category. Then for any $A \in \mathrm{Obj}(\mathscr{C})$, the functors h^A and h_A may be considered as functors to **Ab**, rather than **Set**. The resulting functors are additive.

LEMMA 8.8.11. *A functor $F\colon \mathscr{C} \to \mathscr{D}$ of additive categories is additive if and only if F preserves biproducts, which is to say that the natural morphisms $F(A_1) \oplus F(A_2) \to F(A_1 \oplus A_2)$ and $F(A_1 \oplus A_2) \to F(A_1) \oplus F(A_2)$ are inverse isomorphisms for all objects A_1, A_2 in $\mathscr{C}$.*

PROOF. Suppose first that F is an additive functor. Note that $F(\iota_i \circ p_i) = \mathrm{id}_{F(A_i)}$ and $F(\iota_i \circ p_j) = F(0)$ for $i \neq j$, but $F(0) = 0$ by additivity of F. Again by additivity of F, we have

$$F(\iota_1) \circ F(p_1) + F(\iota_2) \circ F(p_2) = F(\mathrm{id}_{A_1 \oplus A_2}) = \mathrm{id}_{F(A_1 \oplus A_2)}.$$

It follows that $F(A_1 \oplus A_2)$ is a biproduct of $F(A_1)$ and $F(A_2)$ in $\mathscr{D}$, so in particular it is a coproduct.

On the other hand, if F preserves biproducts and $f, g\colon A \to B$ are morphisms in $\mathscr{C}$, then it is easy to see that $F(f \oplus g) = F(f) \oplus F(g)$, and Lemma 8.8.8 tells us that

$$F(f + g) = F(\nabla_B \circ (f \oplus g) \circ \Delta_A) = \nabla_{F(B)} \circ (F(f) \oplus F(g)) \circ \Delta_{F(A)} = F(f) + F(g).$$

□

COROLLARY 8.8.12. *Let $F\colon \mathscr{C} \to \mathscr{D}$ be a fully faithful functor of additive categories. Then F is an additive functor.*

REMARK 8.8.13. For additive functors, we may consider a finer notion of representability than we previously studied. If $F\colon \mathscr{C} \to \mathbf{Ab}$ is an additive contravariant (resp., covariant) functor of additive categories, then we may consider it to be representable if there exists an object $X \in \mathrm{Obj}(\mathscr{C})$ and a natural isomorphism $\eta\colon h^X \rightsquigarrow F$ (resp., $\eta\colon h_X \rightsquigarrow F$). In this case, the morphisms η_A for $A \in \mathrm{Obj}(\mathscr{C})$ will be isomorphisms of groups.

LEMMA 8.8.14. *A morphism in an additive category that admits kernels is a monomorphism if and only if it has zero kernel. A morphism in an additive category that admits cokernels is an epimorphism if and only if it has zero cokernel.*

PROOF. Let $f\colon A \to B$ be a monomorphism, and let $h\colon \ker f \to A$ be the induced morphism. Since $f \circ h = 0$ by definition of the kernel, we have $h = 0$, as f is a monomorphism. This forces $\ker f$ to be 0, since h factors through 0. (Or, one could just apply Lemma 8.7.3.) On the other hand, suppose that f has trivial kernel, and let $g, h\colon C \to A$ be maps with $f \circ g = f \circ h$. Then $f \circ (g - h) = 0$, and by universal property of the kernel, $g - h$ factors through 0, i.e., is 0.

The proof for cokernels is similar, or is the result on kernels in the opposite (additive) category. □

PROPOSITION 8.8.15. *Let $\mathscr{C}$ be an additive category that admits kernels and cokernels. Let $f\colon A \to B$ be a morphism in $\mathscr{C}$. Then*

$$\mathrm{im}\, f \cong \ker(B \to \mathrm{coker}\, f)$$

and

$$\mathrm{coim}\, f \cong \mathrm{coker}(\ker f \to A).$$

PROOF. We prove the first isomorphism. Let $g\colon B \to \mathrm{coker}\, f$. By Yoneda's lemma, it suffices to show that $h^{\mathrm{im}\, f}$ and $h^{\ker g}$ are naturally isomorphic. For $C \in \mathrm{Obj}(\mathscr{C})$, we have a map

$$\mathrm{Hom}_{\mathscr{C}}(C, \mathrm{im}\, f) \xrightarrow{\sim} \{\alpha\colon C \to B \mid \iota_1 \circ \alpha = \iota_2 \circ \alpha\}$$

that takes a morphism $C \to \mathrm{im}\, f$ and composes it with the morphism $\mathrm{im}\, f \to B$ given by definition of the equalizer of the maps $\iota_i\colon B \to B \amalg_A B$. It is a bijection by the universal property of the equalizer.

For any $D \in \mathrm{Obj}(\mathscr{C})$ and morphisms $\phi_1, \phi_2\colon B \to D$ such that $\phi_1 \circ f = \phi_2 \circ f$, note that there exists a unique morphism $k\colon B \amalg_A B \to D$ with $\phi_i = k \circ \iota_i$. Any $\alpha\colon C \to B$ such that $\iota_1 \circ \alpha = \iota_2 \circ \alpha$ then satisfies $\phi_1 \circ \alpha = \phi_2 \circ \alpha$ for any such $\phi_i\colon B \to D$ and any D. On the other hand, note that the ι_i themselves satisfy the property that $\iota_1 \circ f = \iota_2 \circ f$ and are morphisms $\iota_i\colon B \to D$ with $D = B \amalg_A B$. In other words, we have

$$\begin{aligned}&\{\alpha\colon C \to B \mid \iota_1 \circ \alpha = \iota_2 \circ \alpha\}\\ &= \{\alpha\colon C \to B \mid \phi_1 \circ \alpha = \phi_2 \circ \alpha \text{ if } \phi_1 \circ f = \phi_2 \circ f \text{ for some } \phi_1, \phi_2\colon B \to D \text{ (for some } D \in \mathrm{Obj}(\mathscr{C}))\}.\end{aligned}$$

Now, we are in an additive category, so this equals

$$\{\alpha\colon C \to B \mid \phi \circ \alpha = 0 \text{ if } \phi \circ f = 0 \text{ for some } \phi\colon B \to D\}. \tag{8.8.1}$$

By the universal property of $\mathrm{coker}\, f$, for any $\phi\colon B \to D$ with $\phi \circ f = 0$, there is a morphism $j\colon \mathrm{coker}\, f \to D$ with $j \circ g = \phi$. If $g \circ \alpha = 0$, then $\phi \circ \alpha = j \circ g \circ \alpha = 0$, and this works for any

$\phi\colon B \to D$ with $\phi \circ f = 0$. On the other hand, g itself satisfies $g \circ f = 0$ so is such a ϕ. It follows that the set in (8.8.1) equals

$$\{\alpha\colon C \to B \mid g \circ \alpha = 0\}.$$

By the universal property of $\ker g$, this is in bijection with $\mathrm{Hom}_{\mathscr{C}}(C, \ker g)$, taking an α in the set to the unique morphism to $\ker g$ through which it factors. Clearly, the composition of these bijections is natural in C, so we have the desired natural isomorphism. □

DEFINITION 8.8.16. An *abelian* category is an additive category $\mathscr{C}$ in which

i. every morphism in $\mathscr{C}$ admits a kernel and a cokernel and

ii. every morphism in $\mathscr{C}$ is strict.

EXAMPLES 8.8.17.

a. The category R-**mod** is abelian.

b. The full subcategory $\mathscr{C}$ of R-**mod** of finitely generated R-submodules is not necessarily abelian. E.g., when R is commutative and non-Noetherian, we can take I to be an ideal of R that is not finitely generated, and so the kernel of $R \to R/I$ is not in $\mathscr{C}$.

REMARK 8.8.18. Note that if $\mathscr{C}$ is an abelian category, then so is $\mathscr{C}^{\mathrm{op}}$. The roles of mono- and epimorphisms, kernels and cokernels, and images and coimages switch in $\mathscr{C}$ and $\mathscr{C}^{\mathrm{op}}$.

PROPOSITION 8.8.19. *The functor category* $\mathbf{Func}(\mathscr{C}, \mathscr{D})$ *from a small category* $\mathscr{C}$ *to an abelian category* $\mathscr{D}$ *is abelian.*

PROOF. We sketch the proof. First, note that it is additive: we have the zero functor which sends all objects to the zero object and all morphisms to the zero (identity) morphism of the zero object, and if $F, G\colon \mathscr{C} \to \mathscr{D}$ are functors, then $F \oplus G$ is given by $(F \oplus G)(C) = F(C) \oplus G(C)$ and $(F \oplus G)(f) = F(f) \oplus G(f)$ for $f\colon A \to B$ in C. This can be used to define the addition on morphisms (i.e., natural transformations) as before.

Next, the kernel of a natural transformation $\eta\colon F \rightsquigarrow G$ is defined by $(\ker \eta)(C) = \ker \eta_C$ and $(\ker \eta)(f)$ for $f\colon A \to B$ is the kernel of the induced morphism $\ker \eta_A \to \ker \eta_B$. The cokernel is defined similarly. Note that

$$(\mathrm{coim}\, \eta)_A \cong \mathrm{coker}(\ker \eta \rightsquigarrow \eta)_A \cong \mathrm{coker}(\ker \eta_A \rightsquigarrow \eta_A) \cong \mathrm{coim}\, \eta_A,$$

and similarly for images. Finally, since $\mathscr{D}$ is abelian, the natural map $\mathrm{coim}\, \eta \rightsquigarrow \mathrm{im}\, \eta$ is an isomorphism $\mathrm{coim}\, \eta_A \to \mathrm{im}\, \eta_A$ on objects A in $\mathscr{C}$, hence has a natural inverse determined by the inverses of these morphisms. □

TERMINOLOGY 8.8.20. In an abelian category $\mathscr{C}$, we typically refer to a coproduct (when it exists) as a direct sum, and we write $\bigoplus_{i \in I} A_i$ in place of $\coprod_{i \in I} A_i$.

CHAPTER 9

Module theory

9.1. Associative algebras

Much as with groups, we may speaker of the center of a ring.

DEFINITION 9.1.1. The *center* $Z(R)$ of a ring R is the subring of R given by the subset

$$Z(R) = \{a \in R \mid ab = ba \text{ for all } b \in R\}.$$

We now define the notion of an algebra over a commutative ring.

DEFINITION 9.1.2. Let R be a commutative ring. An (associative) *R-algebra* A is the pair of an R-module A and a binary operation $\cdot$ on A which, together with the addition on A, makes A into a ring, and which satisfies

$$r(a \cdot b) = (ra) \cdot b = a \cdot (rb)$$

for all $r \in R$ and $a, b \in A$.

REMARK 9.1.3. An R-algebra A comes endowed with a homomorphism $\phi \colon R \to Z(A)$, given by $\phi(r) = r \cdot 1$ for $r \in R$ and the element $1 \in A$. In fact, to give an R-algebra A is to give a ring A and a ring homomorphism $\phi \colon R \to Z(A)$ with $\phi(1) = 1$, for then this provides an R-module structure on A given by $r \cdot a = \phi(r)a$, which makes A into an R-algebra.

REMARK 9.1.4. Often, it is supposed that the map $\phi \colon R \to Z(A)$ that defines the R-algebra structure on A is injective. This is automatically the case if R is a field.

EXAMPLES 9.1.5. Let R be a commutative ring.

i. The polynomial ring $R[x_1, \ldots, x_n]$ is an R-algebra for any $n \geq 1$.

ii. The matrix ring $M_n(R)$ is an R-algebra for any $n \geq 1$.

iii. If F is a field and E is a field extension of F, then F is an E-algebra.

iv. The ring R is a $\mathbb{Z}$-algebra.

EXAMPLE 9.1.6. The ring $\mathbb{H} = \{a + bi + cj + dk \mid a, b, c, d \in \mathbb{R}\}$ of quaternions has center $Z(\mathbb{H}) = \mathbb{R}$, hence is an algebra over $\mathbb{R}$. Note that $\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\}$ is contained in $\mathbb{H}$, but $\mathbb{H}$ is not a $\mathbb{C}$-algebra, as $\mathbb{C}$ is not contained in the center of $\mathbb{H}$.

DEFINITION 9.1.7. An *R-algebra homomorphism* $f \colon A \to B$ is a ring homomorphism of R-algebras A and B that is also a homomorphism of R-modules.

EXAMPLE 9.1.8. For any R-algebra A, the structure homomorphism $\phi \colon R \to A$ with image in the $Z(A)$ is a homomorphism of R-algebras.

We provide three other classes of examples.

DEFINITION 9.1.9. Let R be a commutative ring with unity, and let X be a set. The *free R-algebra $R\langle X\rangle$* on X is the R-algebra with underlying additive group the free R-module on the words in X and multiplication the unique R-bilinear map given on words in X by concatenation.

REMARK 9.1.10. Another way of describing the free R-algebra on a set X is that it's a noncommutative polynomial ring with variables in X.

NOTATION 9.1.11. If $X = \{x_1, \dots, x_n\}$ has n elements, then we write $R\langle x_1, \dots, x_n\rangle$ for $R\langle X\rangle$.

DEFINITION 9.1.12. Let R be a commutative ring (with unity) and M an R-module. Then endomorphism ring $\mathrm{End}_R(M)$ of M over R is the R-algebra of R-linear endomorphisms of M. It is a ring under addition and composition of endomorphisms, and has the R-module structure given by multiplication of scalars: that is,

$$(r \cdot f)(m) = r \cdot f(m)$$

for all $f \in \mathrm{End}_R(M)$, $r \in R$, and $m \in M$.

REMARK 9.1.13. The map $\phi \colon R \to Z(\mathrm{End}_R(M))$ defining the R-algebra structure on $\mathrm{End}_R(M)$ takes $r \in R$ to left multiplication by r on M.

DEFINITION 9.1.14. The *automorphism group* $\mathrm{Aut}_R(M)$ of an R-module M is the group of R-automorphisms of M under composition.

REMARK 9.1.15. The unit group of $\mathrm{End}_R(M)$ is $\mathrm{Aut}_R(M)$.

EXAMPLE 9.1.16. We have an isomorphism of R-algebras $\mathrm{End}_R(R^n) \xrightarrow{\sim} M_n(R)$ by taking $\phi \in \mathrm{End}_R(R^n)$ to the matrix A defined by $A \cdot e_j = \sum_{i=1}^n a_{ij} e_i$ for the standard basis $\{e_1, \dots, e_n\}$ of R^n. We have that $\phi \in \mathrm{Aut}_R(R^n)$ if and only if A is invertible.

PROPOSITION 9.1.17. *Let M be an R-module and A be an R-algebra. There is a bijection between operations $\cdot \colon A \times M \to M$ which make M into a left A-module and R-algebra homomorphisms $\psi \colon A \to \mathrm{End}_R(M)$ determined by $\psi(a)(m) = a \cdot m$ for all $a \in A$ and $m \in M$.*

We turn to another class of algebras known as group rings. The reader will easily check the following.

LEMMA 9.1.18. *Let R be a commutative ring an G be a group. The set $R[G]$ of elements $\sum_{g \in G} a_g g$ with $a_g \in R$ for all $g \in G$ and almost all $a_g = 0$ and addition and multiplication of multiplication are given respectively by the formulas*

$$\sum_{g \in G} a_g g + \sum_{g \in G} b_g g = \sum_{g \in G} (a_g + b_g) g \quad \text{and} \quad \left(\sum_{g \in G} a_g g\right) \cdot \left(\sum_{g \in G} b_g g\right) = \sum_{g \in G} \left(\sum_{h \in G} a_h b_{h^{-1}g}\right) g$$

is an R-algebra with R-module structure given by the scalar multiplication

$$r \cdot \sum_{g \in G} a_g g = \sum_{g \in G} (r a_g) g.$$

REMARK 9.1.19. As an R-module, $R[G]$ is simply the module $\sum_{g \in G} Rg$. The multiplication on $R[G]$ is the unique multiplication that restricts to the multiplication on G and makes $R[G]$ into an R-algebra. The identity element 1 in $R[G]$ is the identity of G.

DEFINITION 9.1.20. The *group ring* $R[G]$ of a group G with coefficients in a commutative ring R is the unique R-algebra that is free as an R-module with basis G and has multiplication that restricts to the multiplication on G.

EXAMPLE 9.1.21. For $n \geq 1$, there is an isomorphism

$$R[x]/(x^n - 1) \xrightarrow{\sim} R[\mathbb{Z}/n\mathbb{Z}], \qquad \sum_{i=0}^{n-1} a_i x^i \mapsto \sum_{i=0}^{n-1} a_i[i],$$

of R-modules, where $[i]$ denotes the group element corresponding to $i \in \mathbb{Z}/n\mathbb{Z}$. Similarly, one has $R[x, x^{-1}] \cong R[\mathbb{Z}]$.

9.2. Homomorphism groups

REMARK 9.2.1. Let M be a left module over an R-algebra A. Then M is an R-module under $r \cdot m = \phi(r)m$, where $\phi\colon R \to Z(A)$ is given by the structure of A as an R-algebra.

DEFINITION 9.2.2. Let M and N be left modules over an R-algebra A. The *homomorphism group* $\mathrm{Hom}_A(M,N)$ is the R-module of homomorphisms $\phi\colon M \to N$ under the usual addition of maps and the scalar multiplication $(r \cdot \phi)(m) = r \cdot \phi(m)$ for $r \in R$ and $m \in M$.

REMARK 9.2.3. It is traditional to call $\mathrm{Hom}_A(M,N)$ a homomorphism group, even when it has an additional R-module structure (for when we simply take $R = \mathbb{Z}$, it is just a $\mathbb{Z}$-module, or abelian group).

EXAMPLE 9.2.4. Let R be a commutative ring. Then $\mathrm{Hom}_R(R^m, R^n)$ is a free R-module of rank mn, isomorphic to $M_{nm}(R)$ via $\phi \mapsto A$ with $\phi(e_j) = \sum_{i=1}^n a_{ij} e_j$.

EXAMPLE 9.2.5. Let $m, n \geq 1$. Then $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/m\mathbb{Z}, \mathbb{Z}/n\mathbb{Z}) \cong \mathbb{Z}/(m,n)\mathbb{Z}$. That is, an element this group is completely determined of $\phi(1)$, and $\phi(1)$ has to be an element of order dividing m in $\mathbb{Z}/n\mathbb{Z}$, so a multiple of $\frac{n}{\gcd(m,n)}$.

In general, if M and N are A-modules with an additional right module structures that turn them into bimodules, then we can consider transfer these structures to $\mathrm{Hom}_A(M,N)$, as we briefly explore.

DEFINITION 9.2.6. Let A and B be algebras over a commutative ring R. We say that an A-B-bimodule M is *R-balanced* if $rm = mr$ for all $r \in R$ and $m \in M$.

EXAMPLES 9.2.7.

a. If A is an R-algebra, then A is an R-balanced A-A-bimodule.

b. For a commutative ring R, the $M_m(R)$-$M_n(R)$-bimodule $M_{mn}(R)$ is R-balanced.

PROPOSITION 9.2.8. *Let A, B, and C be R-algebras, let M be an R-balanced A-B-bimodule, and let N be an R-balanced A-C-bimodule. Then* $\mathrm{Hom}_A(M,N)$ *is an R-balanced B-C-bimodule under the actions given by*

$$(b\cdot\phi)(m)=\phi(mb) \quad \text{and} \quad (\phi\cdot c)(m)=\phi(m)c$$

for $b\in B$, $c\in C$, $m\in M$, and $\phi\in\mathrm{Hom}_A(M,N)$.

Homomorphism groups behave well with respect to direct sums and products, as made precise in the following proposition.

PROPOSITION 9.2.9. *Let A be an R-algebra.*

a. Let M be a left A-module, and let $\{N_j \mid j\in J\}$ be a collection of left A-modules. Then there is a canonical isomorphism of left R-modules

$$\mathrm{Hom}_A\Big(M,\prod_{j\in J}N_j\Big)\cong\prod_{j\in J}\mathrm{Hom}_A(M,N_j).$$

b. Let N be a left A-module, and let $\{M_i \mid i\in I\}$ be a collection of left A-modules. Then there is a canonical isomorphism of left R-modules

$$\mathrm{Hom}_A\Big(\bigoplus_{i\in I}M_i,N\Big)\cong\prod_{i\in I}\mathrm{Hom}_A(M_i,N).$$

PROOF. Given a collection of A-module homomorphisms $\phi_j\colon M\to N_j$ for $j\in J$, we define $\Phi\colon M\to\prod_{j\in J}N_j$ by $\Phi(m)=(\phi_j(m))_{j\in J}$, which is clearly an A-module homomorphism. Conversely, given Φ, we define $\phi_j=\pi_j\circ\Phi$ where $\pi_j\colon\prod_{j\in J}N_j$ is the projection map, and ϕ_j is then an A-module homomorphism. The bijection $(\phi_j)_{j\in J}\mapsto\Phi$ is clearly a map of R-modules. Thus, we have part a.

Now, given a collection of A-module homomorphisms $\psi_i\colon M_i\to N$ for $i\in I$, we define $\Psi\colon\bigoplus_{i\in I}M_i\to N$ by $\Psi((m_i)_{i\in I})=\sum_{i\in I}m_i$, which is well-defined as all but finitely many $m_i=0$ by definition of the direct sum. The map Ψ is then an A-module homomorphism. Conversely, given Ψ, we define $\psi_i(m)=\Psi(\iota_i(m))$, where $\iota_i\colon M\to\bigoplus_{i\in I}M_i$ is the inclusion. These are by definition inverse associations, and the bijection $(\psi_i)_{i\in I}\to\Psi$ is again clearly an R-module homomorphism. □

Let us consider the example of a dual vector space.

DEFINITION 9.2.10. Let V be a vector space over a field K. The *dual vector space* is $V^*=\mathrm{Hom}_K(V,K)$.

REMARK 9.2.11. Note that $V\cong\bigoplus_{i\in I}K$ for any choice of basis, so

$$V^*\cong\mathrm{Hom}_K\Big(\bigoplus_{i\in I}K,K\Big)\cong\prod_{i\in I}\mathrm{Hom}_K(K,K)\cong\prod_{i\in I}K$$

by part b of Proposition 9.2.9. That is, V and V^* are not in general isomorphic, but they will be so if V is finite-dimensional. However, this isomorphism is not canonical: it depends on a choice of basis, which we next make explicit.

DEFINITION 9.2.12. Let V be an n-dimensional vector space over a field K, and let $B = \{e_1, e_2, \dots, e_n\}$ be a basis of V. The dual basis to B is the basis of V^* given by $B^* = \{f_1, f_2, \dots, f_n\}$, where for $1 \le i, j \le n$, we have

$$f_i(e_j) = \delta_{ij}$$

We next consider the double dual $V^{**} = (V^*)^*$ of an arbitrary vector space. For a finite-dimensional vector space, it is canonically isomorphic to V.

PROPOSITION 9.2.13. *Let V be a vector space over a field K. There is a canonical injection $\Phi\colon V \to V^{**}$ of K-vector spaces given by $F(v)(f) = f(v)$ for $v \in V$ and $f \in V^*$. It is an isomorphism if V is finite-dimensional.*

PROOF. Let $v \in V$ and $f \in V^*$. First, note that

$$\Phi(v)(f + af') = f(v) + af(v') = \Phi(v)(f) + a\Phi(v)(f'),$$

so $\Phi(v) \in V^{**}$. Second, note that

$$\Phi(av + v')(f) = f(av + v') = af(v) + f(v') = a \cdot \Phi(v)(f) + \Phi(v')(f),$$

so Φ is a K-linear transformation. Third, note that if $\Phi(v) = 0$, then $\Phi(v)(f) = f(v) = 0$ for all $f \in V^*$. If $v \neq 0$, we can extend $\{v\}$ to a basis B of V and define $f \in V^*$ by $f(v) = 1$ and $f(w) = 0$ for all $w \in B - \{v\}$. Thus, the fact that $f(v) = 0$ for all $f \in V^*$ implies that $v = 0$, so Φ is injective.

Now, suppose that V is n-dimensional, let $\{e_1, e_2, \dots, e_n\}$ be a basis, and let $\{f_1, f_2, \dots, f_n\}$ be its dual basis in V^*. If $\varphi \in V^{**}$, then set $c_j = \varphi(f_j)$ for each $1 \le j \le n$. Then

$$\Phi\Big(\sum_{i=1}^{n} \varphi(f_i)e_i\Big)(f_j) = f_j\Big(\sum_{i=1}^{n} \varphi(f_i)e_i\Big) = \varphi(f_j)$$

for all j, so $\varphi \in \Phi(V)$. That is, Φ is an isomorphism. □

9.3. Tensor products

DEFINITION 9.3.1. Let A be ring, let M be a right A-module, and let N be a left A-module. The *tensor product* $M \otimes_A N$ of M and N over A is the abelian group that is the quotient of the free abelian group with basis $M \times N$ by its subgroup generated by

i. $(m + m', n) - (m, n) - (m', n)$ for all $m, m' \in M$ and $n \in N$,

ii. $(m, n + n') - (m, n) - (m, n')$ for all $m \in M$ and $n, n' \in N$, and

iii. $(ma, n) - (m, an)$ for all $m \in M$, $n \in N$, and $a \in A$.

The image of (m, n) in $M \otimes_A N$ is denoted $m \otimes n$.

DEFINITION 9.3.2. Let A be ring, let M be a right A-module, and let N be a left A-module. An element of $M \otimes_A N$ of the form $m \otimes n$ for some $m \in M$ and $n \in N$ is called a *simple tensor*.

REMARK 9.3.3. Any tensor product $M \otimes_A N$ is generated as an abelian group by simple tensors $m \otimes n$ for $m \in M$ and $n \in N$. It is not in general equal to the set of such tensors.

PROPOSITION 9.3.4. *Let A be an algebra over a commutative ring R, let M be a right A-module, and let N be a left A-module. The tensor product $M \otimes_A N$ is an R-module under the unique action that satisfies*

$$r(m \otimes n) = mr \otimes n = m \otimes rn.$$

for all $r \in R$, $m \in M$, and $n \in N$.

PROOF. If we consider the free abelian group on $M \times N$ as an R-module via $r(m,n) = (mr,n)$ for $r \in R$, $m \in M$, and $n \in N$, then the elements providing the relations in Definition 9.3.1 define an R-submodule. Therefore, the quotient becomes an R-module under this action. □

REMARK 9.3.5. If A is an R-algebra, the tensor product $M \otimes_A N$ is isomorphic to the quotient of the free R-module on $M \times N$ by the submodule generated by the elements of Definition 9.3.1, along with the elements $r(m,n) - (mr,n)$ for $r \in R$, $m \in M$, and $n \in N$.

DEFINITION 9.3.6.

a. Let L, M, and N be abelian groups. A map $\phi \colon M \times N \to L$ is said to be *bilinear* if

$$\phi(m+m',n) = \phi(m,n) + \phi(m',n) \quad \text{and} \quad \phi(m,n+n') = \phi(m,n) + \phi(m,n').$$

for all $m,m' \in M$ and $n,n' \in N$. Here, the first equality (for all m, m', and n) is referred to as left linearity (or linearity in the first variable) and the second as right linearity.

b. Let L, M, and N be left modules over a commutative ring R. A bilinear map $\phi \colon M \times N \to L$ satisfying

$$r\phi(m,n) = \phi(rm,n) = \phi(m,rn)$$

for all $r \in R$, $m \in M$, and $n \in N$, then ϕ is said to be *R-bilinear*.

DEFINITION 9.3.7. Let A be a ring, let M be a right A-module, and let N be a left A-module. A function $\phi \colon M \times N \to L$ is said to be *A-balanced* if $\phi(ma,n) = \phi(m,an)$ for all $a \in A$.

REMARK 9.3.8. Let A be an algebra over a commutative ring R, let M be a right A-module, and let N be a left A-module. The tensor product $M \otimes_A N$ is endowed with an A-balanced R-bilinear map

$$\iota_{M,N} \colon M \times N \to M \otimes_A N, \qquad \phi(m,n) = m \otimes n,$$

as seen directly from the relations defining $M \otimes_A N$.

The tensor product enjoys a universal property, exhibited in the following proposition.

PROPOSITION 9.3.9. *Let A be an algebra over a commutative ring R, let M be a right A-module, and let N be a left A-module. Let $\phi \colon M \times N \to L$ be R-bilinear and A-balanced. Then there exists a unique R-module homomorphism $\Phi \colon M \otimes_A N \to L$ such that $\Phi(m \otimes n) = \phi(m,n)$ for all $m \in M$ and $n \in N$.*

PROOF. We use the alternate construction of $M \otimes_A N$ of Remark 9.3.5. The map ϕ induces a unique R-module homomorphism

$$F \colon \bigoplus_{(m,n) \in M \times N} R(m,n) \to L, \qquad F((m,n)) = \phi(m,n),$$

since the direct sum is free. The R-bilinearity of ϕ tells us that the elements $(m+m',n)-(m,n)-(m',n)$, $(m,n+n')-(m,n)-(m,n')$, and $r(m,n)-(mr,n)$ lie its the kernel. The fact that ϕ is A-balanced similarly tells us that the elements $(ma,n)-(m,an)$ are contained in its kernel. The first homomorphism theorem then provides an R-module homomorphism $\Phi\colon M\otimes_A N\to L$ with $\Phi(m\otimes n)=\phi(m,n)$ for all $m\in M$ and $n\in N$.

If $\Psi\colon M\otimes_A N\to L$ is an R-module homomorphism also satisfying $\Psi\circ\iota_{M,N}=\phi$, then $\Psi(m\otimes n)=\phi(m,n)=\Phi(m\otimes n)$ for all $m\in M$ and $n\in N$, but the symbols $m\otimes n$ generate $M\otimes_A N$ as an R-module, since the tensor product is defined as the quotient of the free R-module on $M\times N$. Therefore, we must have $\Phi=\Psi$. □

REMARK 9.3.10. The defining property of the map $\Phi\colon M\otimes_A N\to L$ of Proposition 9.3.9 is stated more succinctly as $\Phi\circ\iota_{M,N}=\phi$.

The reader may check the following, which gives the uniqueness of the tensor product up to unique isomorphism as a module satisfying the universal property of the tensor product.

PROPOSITION 9.3.11. *Let A be an algebra over a commutative ring R, let M be a right A-module, and let N be a left A-module. Let P be an R-module, and let $\lambda: M\times N\to P$ be an R-bilinear map such that for any R-bilinear, A-balanced map $\phi: M\times N\to L$, there exists a unique R-module homomorphism $\Phi\colon P\to L$ such that $\Phi\circ\lambda=\phi$. Then there is a unique isomorphism $\psi\colon P\xrightarrow{\sim} M\otimes_A N$ such that $\psi\circ\lambda=\iota_{M,N}$.*

REMARK 9.3.12. Let A be an algebra over a commutative ring R, let M be a right A-module, and let N be a left A-module. For any $m\in M$ and $n\in N$, we have $m\otimes 0=0=0\otimes n$. For the first equality, note that $m\otimes 0=0(m\otimes 0)=0$.

We give an example by way of a proposition.

PROPOSITION 9.3.13. *Let $m,n\geq 1$. Then $(\mathbb{Z}/m\mathbb{Z})\otimes_{\mathbb{Z}}(\mathbb{Z}/n\mathbb{Z})\cong\mathbb{Z}/(m,n)\mathbb{Z}$.*

PROOF. Let $d=\gcd(m,n)$, and write $d=am+bn$ for some $a,b\in\mathbb{Z}$. Note that $x\otimes y=xy(1\otimes 1)$, so $T=(\mathbb{Z}/m\mathbb{Z})\otimes_{\mathbb{Z}}(\mathbb{Z}/n\mathbb{Z})$ is cyclic, and moreover,

$$d(1\otimes 1)=(am+bn)(1\otimes 1)=a(m\otimes 1)+b(1\otimes n)=0,$$

so the order of T divides d.

We can define a bilinear map $\phi: \mathbb{Z}/m\mathbb{Z}\times\mathbb{Z}/n\mathbb{Z}\to\mathbb{Z}/d\mathbb{Z}$ by $\phi(x,y)=xy \bmod d$ for $x\in\mathbb{Z}/m\mathbb{Z}$ and $y\in\mathbb{Z}/n\mathbb{Z}$. We then have a homomorphism $\Phi\colon T\to\mathbb{Z}/(m,n)\mathbb{Z}$ with $\Phi(1\otimes 1)=1$, and it is therefore surjective. This forces $|T|=d$ and Φ to be an isomorphism, as desired. □

PROPOSITION 9.3.14. *Let A be an R-algebra, let M be a right A-module, and let $\{N_i \mid i\in I\}$ be a collection of left A-modules. Then*

$$M\otimes_A\Big(\bigoplus_{i\in I}N_i\Big)\cong\bigoplus_{i\in I}(M\otimes_A N_i).$$

PROOF. First, define an R-bilinear, A-balanced map

$$\phi: M\times\Big(\bigoplus_{i\in I}N_i\Big)\to\bigoplus_{i\in I}(M\otimes_A N_i),\qquad \phi(m,\sum_{i\in I}n_i)=\sum_{i\in I}m\otimes n_i.$$

This induces an R-module homomorphism

$$\Phi\colon M\otimes_A\Big(\bigoplus_{i\in I}N_i\Big)\cong\bigoplus_{i\in I}(M\otimes_A N_i)$$

with $\Phi(m\otimes n_i)=m\otimes n_i$ for $m\in M$ and $n_i\in I$ for some $i\in I$.

Next, define R-bilinear, A-balanced maps

$$\psi_i\colon M\times N_i\to M\otimes_A\Big(\bigoplus_{i\in I}N_i\Big),\qquad \psi_i(m,n_i)=m\otimes n_i.$$

The collection $(\psi_i)_{i\in I}$ gives rise to a unique R-module homomorphism

$$\Psi\colon\bigoplus_{i\in I}(M\otimes_A N_i)\to M\otimes_A\Big(\bigoplus_{i\in I}N_i\Big),$$

satisfying $\Psi(m\otimes n_i)=m\otimes n_i$ for m and n_i as above by Proposition 9.2.9b. By definition, the maps Φ and Ψ are inverse to each other. □

PROPOSITION 9.3.15. *Let M and N be modules over a commutative ring R. Then there is a unique isomorphism of R-modules*

$$M\otimes_R N\xrightarrow{\sim}N\otimes_R M,\qquad m\otimes n\mapsto n\otimes m.$$

PROOF. Consider the R-bilinear map $\phi\colon M\times N\to N\otimes_R M$ given by $\phi(m,n)=n\otimes m$. It induces an R-module homomorphism $\Phi\colon M\otimes_R N\to N\otimes_R M$ satisfying $\Phi(m\otimes n)=n\otimes m$ by the universal property of the tensor product. It then has inverse the similarly defined map $\Psi\colon N\otimes_R M\to M\otimes_R N$ with $\Psi(n\otimes m)=m\otimes n$. □

REMARK 9.3.16. We can allow a tensor product over an arbitrary R-algebra A in Proposition 9.3.15, but we obtain $M\otimes_A N\cong N\otimes_{A^{\mathrm{op}}}M$ as R-modules (noting that A^{op} has the same R-module structure as A).

EXAMPLE 9.3.17. Let R be a commutative ring. The tensor product $R^m\otimes_R R^n$ is a free R-module of rank mn with basis $\{e_i\otimes e_j\mid 1\le i\le m, 1\le j\le n\}$. This follows immediately from Proposition 9.3.14 (and Proposition 9.3.15) and the fact that $R\otimes_R R\cong R$. Here, the latter isomorphism is induced the R-bilinear map $(x,y)\mapsto xy$, its inverse being the map $R\to R\otimes_R R$ with $x\mapsto x\otimes 1$.

PROPOSITION 9.3.18. *Let A and B be R-algebras. Let L be a right A-module, let M be an R-balanced A-B-bimodule, and let N be a left B-module. Then there is a unique isomorphism of R-modules*

$$(L\otimes_A M)\otimes_B N\xrightarrow{\sim}L\otimes_A(M\otimes_B N),\qquad (l\otimes m)\otimes n\mapsto l\otimes(m\otimes n).$$

PROOF. Let $\phi\colon(L\otimes_A M)\times N\to P$ be an R-bilinear, B-balanced map to some R-module P. This gives rise to a map

$$\psi=\phi\circ(\iota_{L,M}\times\mathrm{id}_N)\colon L\times M\times N\to P$$

which is R-linear in each variable separately and satisfies

$$\psi(la,m,n)=\psi(l,am,n)\quad\text{and}\quad\psi(l,mb,n)=\psi(l,m,bn)$$

for all $a \in A$, $b \in B$, $l \in L$, $m \in M$, and $n \in N$. In particular, for each $l \in L$, we obtain an R-module homomorphism $\Psi_l \colon M \otimes_B N \to P$ with $\Psi_l(m \otimes n) = \psi(l,m,n)$ by the fact that $\psi_l \colon M \times N \to P$ with $\psi_l(m,n) = \psi(l,m,n)$ is R-bilinear and B-balanced. We then obtain an R-bilinear, A-balanced map

$$\theta \colon L \times (M \otimes_B N) \to P, \qquad \theta(l, m \otimes n) = \psi_l(m \otimes n) = \psi(l,m,n)$$

which in turn induces an R-module homomorphism

$$\Theta \colon L \otimes_A (M \otimes_B N) \to P, \qquad \Theta(l \otimes (m \otimes n)) = \psi(l,m,n).$$

Since the elements $l \otimes (m \otimes n)$ generate $L \otimes_A (M \otimes_B N)$, this is the unique homomorphism that agrees with ψ on these simple tensors. Since $\psi(l,m,n) = \phi(l \otimes m, n)$, the R-module $L \otimes_A (M \otimes_B N)$ satisfies the universal property of the tensor product $L \otimes_A (M \otimes_B N)$, hence is canonically isomorphic to it via the indicated map, as in Proposition 9.3.11. □

LEMMA 9.3.19. *Let A be an R-algebra. Let M and M' be right A-modules, and let N and N' be left A-bmodules. Let $\phi \colon M \to M'$ and $\psi \colon N \to N'$ be homomorphisms of left and right A-modules, respectively. Then there exists a homomorphism of R-modules*

$$\phi \otimes \psi \colon M \otimes_A N \to M' \otimes_A N', \qquad (\phi \otimes \psi)(m \otimes n) = \phi(m) \otimes \psi(n).$$

PROOF. The map $\theta \colon M \times N \to M' \times N'$ with $\theta(m,n) = \phi(m) \otimes \psi(n)$ is immediately seen to be R-bilinear, and it is A-balanced since

$$\phi(ma) \otimes \psi(n) = \phi(m)a \otimes \psi(n) = \phi(m) \otimes a\psi(n) = \phi(m) \otimes \psi(an).$$

Thus, it induces an R-module homomorphism $M \otimes_A N \to M' \otimes_A N'$ with the desired property. □

We can also form the tensor product of R-algebras.

PROPOSITION 9.3.20. *Let A and B be algebras over a commutative ring R. The tensor product $A \otimes_R B$ is an R-algebra under the unique multiplication satisfying*

$$(a \otimes b)(a' \otimes b') = aa' \otimes bb'.$$

for $a, a' \in A$ and $b, b' \in B$.

PROOF. First, we should check the desired multiplication on $A \otimes_R B$ is well-defined. To start, given $a \in A$ and $b \in B$, we claim that the map $A \times B \to A \otimes_R B$ given by $(a', b') \mapsto aa' \otimes bb'$ is R-bilinear (and therefore R-balanced). To see this, we merely note that $a(ra') \otimes bb' = r(aa' \otimes bb')$ and $a(a' + a'') \otimes bb' = aa' \otimes bb' + aa'' \otimes bb'$. Therefore, we obtain a well-defined map

$$\psi \colon A \times B \to \mathrm{End}_R(A \otimes_R B), \qquad \psi(a,b)(a' \otimes b') = aa' \otimes bb'.$$

Note also that ψ is R-bilinear as well, so we obtain an R-module homomorphism $A \otimes_R B \to \mathrm{End}_R(A \otimes_R B)$, which we may rewrite then as a well-defined operation

$$(A \otimes_R B) \times (A \otimes_R B) \to A \otimes_R B, \qquad (a \otimes b, a' \otimes b') \mapsto aa' \otimes bb'.$$

This operation is R-bilinear by what we have said. As it clearly satisfies $(1 \otimes 1)(a' \otimes b') = a' \otimes b'$, so we need only observe its associativity to finish the proof of the result. This can be checked on simple tensors, for which it is in an immediate consequence of the associativity of the operations on A and B. □

PROPOSITION 9.3.21. *Let A and B be algebras over a commutative ring R. An abelian group M that is a left A-module and a right B-module is an R-balanced A-B-bimodule if and only if it is an $A \otimes_R B^{\mathrm{op}}$-module under the action $(a \otimes b)m = (am)b$.*

PROOF. Let M be an R-balanced A-B-bimodule. We endow it with an $A \times B^{\mathrm{op}}$-action by $(a,b)m = amb$. This is action is R-bilinear, so it factors through an action of $A \otimes_R B^{\mathrm{op}}$ that clearly satisfies $(1 \otimes 1)m = m$ and $(a \otimes b)(m+m') = (a \otimes b)m + (a \otimes b)m'$ and therefore makes M into an $A \otimes_R B^{\mathrm{op}}$-module.

Conversely, if M is an $A \otimes_R B^{\mathrm{op}}$-module, it is in particular an R-balanced A-B-bimodule via the actions $am = (a \otimes 1)m$ and $mb = (1 \otimes b)m$, as the reader may quickly verify. □

When M and N have R-balanced bimodule structures, we can also attain a bimodule structure on their tensor product.

PROPOSITION 9.3.22. *Let A, B, and C be R-algebras over a commutative ring R. Let M be an A-B-bimodule and N be an R-balanced B-C-bimodule. Then $M \otimes_B N$ is an R-balanced A-C-bimodule with respect to actions satisfying*

$$a(m \otimes n) = (am) \otimes n \quad \text{and} \quad (m \otimes n)c = m \otimes (nc)$$

for all $a \in A$, $c \in C$, $m \in M$, and $n \in N$.

PROOF. For $a \in A$ and $c \in C$, we can define an R-bilinear map

$$\phi \colon M \times N \to M \otimes_B N, \qquad (m,n) \mapsto (am) \otimes (nc),$$

noting that

$$\phi_{a,c}(rm + m', n) = (a(mr + m')) \otimes (nc) = r(am \otimes nc) + am' \otimes nc = r\phi(m,n) + \phi(m,n')$$

and similarly for the second variable. We thus have an induced map

$$\Phi_{a,c} \colon M \otimes_B N \to M \otimes_B N, \qquad m \otimes n \mapsto (am) \otimes_B (nc)$$

of R-modules. The map

$$A \times C^{\mathrm{op}} \to \mathrm{End}_R(M \otimes_B N), \qquad (a,c) \to \Phi_{a,c}$$

then defines an R-algebra homomorphism. In other words, this gives $M \otimes_B N$ the structure of a left $A \otimes_R C^{\mathrm{op}}$-module. □

We give an application.

PROPOSITION 9.3.23. *Let A be a ring, let M be a left A-module, and let I be a two-sided ideal of A. Then there is an isomorphism of left A-modules*

$$M/IM \xrightarrow{\sim} A/I \otimes_A M, \qquad m + IM \mapsto 1 \otimes m.$$

PROOF. Note that A/I is an A-A-bimodule, so $A/I \otimes_A M$ has the structure of an A-module by Proposition 9.3.22. In one direction, we can define an A-module homomorphism $\Psi \colon M \to A/I \otimes_A M$ by $\Psi(m) = 1 \otimes m$. In the other, we can define an left A-linear, A-balanced map $\phi \colon A/I \times M \to M/IM$ by $\phi(a+I, m) = am + IM$, which induces an A-module homomorphism $\Phi \colon A/I \otimes_A M \to M/IM$ which is clearly inverse to Ψ. □

We have the following direct corollary.

COROLLARY 9.3.24. *Let A be a ring, and let M be a left A-module. Then $M \cong A \otimes_A M$ as A-modules.*

REMARK 9.3.25. Note that Proposition 9.3.23 requires I to be a two-sided ideal, though the definition of M/IM only requires I to be a left ideal. That is, we need a right A-action on A/I in order to define $A/I \otimes_A M$. We cannot take $M \otimes_A A/I$ either, as M is a left A-module.

Here is an interesting comparison of tensor products and homomorphism groups in the case of vector spaces.

LEMMA 9.3.26. *Let V and W be finite-dimensional vector spaces over a field F. Then we have an F-linear isomorphism*

$$\Psi \colon V^* \otimes_F W \xrightarrow{\sim} \mathrm{Hom}_F(V,W), \qquad \Psi(\phi \otimes w)(v) = \phi(v)w$$

for $\phi \in V^*$, $v \in V$, and $w \in W$.

PROOF. One checks directly that the map $\psi \colon V^* \times W \to \mathrm{Hom}_F(V,W)$ with $\psi(\phi,w)(v) = \phi(v)w$ is F-bilinear, thus induces a map on the tensor product. Let B be a basis of V and C be a basis of W. For each $v \in B$, and $\varphi \in \mathrm{Hom}_F(V,W)$, write

$$\varphi(v) = \sum_{w \in C} a_{v,w} w.$$

Define $\phi_w \in V^*$ for $w \in C$ by $\phi_w(v) = a_{v,w}$ for $v \in B$. We can then define $\Theta \colon \mathrm{Hom}_F(V,W) \to V^* \otimes_F W$ by

$$\Theta(\varphi) = \sum_{w \in C} \phi_w \otimes w.$$

By definition, $\Psi(\Theta(\varphi))(v) = \phi(v)$ and

$$\Theta(\Psi(\phi \otimes w)) = \Theta(v \mapsto \phi(v)w) = \phi \otimes w.$$

□

We will interpret the following as an adjointness of homomorphism and tensor product functors.

THEOREM 9.3.27. *Let A and B be algebras over a commutative ring R. Let M be an R-balanced A-B-bimodule, let N be a left B module, and let L be a left A-module. Then there is an isomorphism of R-modules*

$$\Xi \colon \mathrm{Hom}_A(M \otimes_B N, L) \xrightarrow{\sim} \mathrm{Hom}_B(N, \mathrm{Hom}_A(M,L))$$

given by

$$\Xi(f)(n)(m) = f(m \otimes n)$$

for all $f \in \mathrm{Hom}_A(M \otimes_B N, L)$, $m \in M$ and $n \in N$.

PROOF. First, define Ξ as in the statement of the theorem. Note that

$$\begin{aligned}\Xi(f)(n)(am+m') &= f((am+m')\otimes n) = f(a(m\otimes n)+(m'\otimes n))\\ &= af(m\otimes n)+f(m'\otimes n) = a\Xi(f)(n)(m)+\Xi(f)(n)(m'),\end{aligned}$$

so $\Xi(f)(n)$ is a homomorphism of A-modules. Moreover,

$$\begin{aligned}\Xi(f)(bn'+n'')(m) &= f(m\otimes(bn'+n'')) = f(mb\otimes n')+f(m\otimes n'')\\ &= \Xi(f)(mb)(n')+\Xi(f)(m)(n'') = (b\Xi(f))(m)(n')+\Xi(f)(m)(n''),\end{aligned}$$

so $\Xi(f)$ is a homomorphism of B-modules. Thus, Ξ is well-defined. In addition,

$$\begin{aligned}\Xi(rf)(n)(m) &= (rf)(m\otimes n) = f(r(m\otimes n)) = f(mr\otimes n)\\ &= f(m\otimes rn) = \Xi(f)(rn)(m) = (r\Xi(f))(n)(m)\end{aligned}$$

so $\Xi(rf) = r\Xi(f)$, and since Ξ is also clearly a homomorphism of abelian groups, Ξ is a homomorphism of R-modules.

To finish the proof, we must exhibit an inverse to Ξ. For this, suppose we are given $\lambda \in \mathrm{Hom}_B(N,\mathrm{Hom}_A(M,L))$ and define $\phi\colon M\times N\to L$ by $\phi(m,n) = \lambda(n)(m)$. This map satisfies

$$\begin{aligned}\phi(rm+m',n) &= \lambda(n)(rm+m') = r\lambda(n)(m)+\lambda(n)(m') = r\phi(m,n)+\phi(m',n),\\ \phi(m,rn+n') &= (r\lambda(n))(m)+\lambda(n')(m) = r\lambda(n)(m)+\lambda(n')(m) = r\phi(m,n)+\phi(m,n'),\\ \phi(mb,n) &= \lambda(mb)(n) = (\lambda(m)b)(n) = \lambda(m)(bn) = \phi(m,bn),\\ \phi(am,n) &= \lambda(am)(n) = (a\lambda(m))(n) = a\lambda(m)(n) = a\phi(m,n)\end{aligned}$$

for all $m,m'\in M$, $n,n'\in N$, $r\in R$, $a\in A$, and $b\in B$. Thus, ϕ induces a unique map $\Phi\colon M\otimes_B N\to L$ of A-modules with $\Phi(m\otimes n) = \lambda(n)(m)$. The map $\lambda\to\Phi$ is then by definition inverse to Ξ, which tells us that Ξ is a bijection, hence an isomorphism. □

REMARK 9.3.28. If we suppose in Theorem 9.3.27 that N is an R-balanced B-C-bimodule and L is an R-balanced A-D-bimodule for R-algebras C and D, then the isomorphism Ξ is one of R-balanced C-D-bimodules.

REMARK 9.3.29. Let A and B be algebras over a commutative ring R. Fix an R-balanced A-B-bimodule M. Define a functor $t_M\colon B\text{-}\mathbf{mod}\to A\text{-}\mathbf{mod}$ by $t_M(N) = M\otimes_B N$ on B-modules N and $t_M(g) = \mathrm{id}_M\otimes g$ on B-module homomorphisms $g\colon N\to N'$. Define another functor $h_M\colon A\text{-}\mathbf{mod}\to B\text{-}\mathbf{mod}$ by $h_M(L) = \mathrm{Hom}_A(M,L)$ on A-modules L and $h_M(f)(h) = f\circ h$ for $f\in\mathrm{Hom}_A(L,L')$ and $h\in\mathrm{Hom}_A(M,L)$. Then the isomorphism of Theorem 9.3.27 is the adjunction map

$$\mathrm{Hom}_A(t_M(N),L)\xrightarrow{\sim}\mathrm{Hom}_B(N,h_M(L)).$$

These isomorphisms are natural in N and L, and hence t_M is left adjoint to h_M.

9.4. Exterior powers

In this section, R will denote a commutative ring.

DEFINITION 9.4.1. Let M be an R-module. For a nonnegative integer k, the kth tensor power $M^{\otimes k}$ of M over R is the tensor product $M \otimes_R M \otimes_R \cdots \otimes_R M$ of k copies of M if k is positive and R if $k = 0$.

DEFINITION 9.4.2. Let $M_1, M_2, \ldots, M_k$ and N be R-modules for some $k \geq 1$. A map $f\colon M_1 \times M_2 \times \cdots \times M_k \to N$ is said to be *R-multilinear* if it is R-linear in each of its k variables, which is to say that

$$\begin{aligned} &f(m_1, m_2, \ldots, m_{i-1}, rm_i + m_i', m_{i+1}, \ldots, m_k) \\ &\qquad = f(m_1, m_2, \ldots, m_{i-1}, rm_i, m_{i+1}, \ldots, m_k) + f(m_1, m_2, \ldots, m_{i-1}, m_i', m_{i+1}, \ldots, m_k) \end{aligned}$$

for $r \in R$ and all m_j and $m_j' \in M_j$ for $1 \leq j \leq k$.

The reader will quickly check the following.

PROPOSITION 9.4.3. *Let $M_1, M_2, \ldots, M_k$ and N be R-modules for some $k \geq 1$. For an R-multilinear map $\theta\colon \prod_{i=1}^k M_i \to N$, there exists a unique R-module homomorphism*

$$\Theta\colon M_1 \otimes_R M_2 \otimes_R \cdots \otimes_R M_k \to N$$

such that $\Theta(m_1 \otimes m_2 \otimes \cdots \otimes m_k) = \theta(m_1, m_2, \ldots, m_k)$ for all $m_i \in M_i$ with $1 \leq i \leq k$.

DEFINITION 9.4.4. Let M be a module over a commutative ring R. For a nonnegative integer k, the kth *exterior power* $\bigwedge^k M$ is the quotient of $M^{\otimes k}$ by the R-submodule generated by the elements of the form $m_1 \otimes m_2 \otimes \cdots \otimes m_k$, where $m_i = m_j$ for some $1 \leq i < j \leq k$. The image of a tensor $m_1 \otimes m_2 \otimes \cdots \otimes m_k$ in $\bigwedge^k M$ is denoted $m_1 \wedge m_2 \wedge \cdots \wedge m_k$.

REMARK 9.4.5. The kth exterior power of a module M is often referred to as the wedge product of M with itself k times.

DEFINITION 9.4.6. Let M and N be abelian groups. A multilinear map $f\colon M^k \to N$ is said to be *alternating* if

$$f(m_1, m_2, \ldots, m_k) = 0$$

for any $m_j \in M$ for $1 \leq j \leq k$ such that $m_i = m_{i+1}$ for some $1 \leq i \leq k-1$.

REMARK 9.4.7. There is an alternating, R-bilinear map $\kappa\colon M^k \to \bigwedge^k M$ for any $k \geq 0$ such that $\kappa(m_1, m_2, \ldots, m_k) = m_1 \wedge m_2 \wedge \cdots \wedge m_k$ for all $m_j \in M$ for $1 \leq j \leq k$

PROPOSITION 9.4.8. *Let M and N be R-modules, and let $\psi\colon M^k \to N$ be R-multilinear and alternating. Then there exists a unique R-module homomorphism $\Psi\colon \bigwedge^k M \to N$ such that*

$$\Psi(m_1 \wedge m_2 \wedge \cdots \wedge m_k) = \psi(m_1, m_2, \ldots, m_k)$$

for all $m_i \in M$ for $1 \leq i \leq k$.

PROOF. Since ψ is R-multilinear, there exists by Proposition 9.4.3 a unique R-module homomorphism $\Theta\colon M^{\otimes k} \to N$ with $\Theta(m_1 \otimes m_2 \otimes \cdots \otimes m_k) = \psi(m_1, m_2, \ldots, m_k)$ for all $m_i \in M$ for $1 \leq i \leq k$. If $m_i = m_{i+1}$ for some $1 \leq i \leq k-1$, then

$$\Theta(m_1 \otimes m_2 \otimes \cdots \otimes m_k) = \psi(m_1, m_2, \ldots, m_k) = 0$$

as ψ is alternating, so Θ factors through the desired map $\Psi\colon \bigwedge^k M \to N$.

If $\Psi'\colon \bigwedge^k M \to N$ also has the property of the proposition, then we may compose Ψ with the quotient map $\kappa\colon M^{\otimes k} \to N$ to obtain a map $\Theta' = \Psi \circ \kappa\colon M^{\otimes k} \to N$ that satisfies the universal property of Proposition 9.4.3, hence is equal to Θ. This then forces the equality $\Psi' = \Psi$ for the induced maps on the exterior product. □

We leave the following to the reader.

LEMMA 9.4.9. *Let $\varphi\colon M \to N$ be a homomorphism of R-modules. Then for any $k \geq 0$, there exists a homomorphism $\bigwedge^k \varphi\colon \bigwedge^k M \to \bigwedge^k N$ satisfying*

$$(\bigwedge^k \varphi)(m_1 \wedge m_2 \wedge \cdots \wedge m_k) = \varphi(m_1) \wedge \varphi(m_2) \wedge \cdots \wedge \varphi(m_k).$$

LEMMA 9.4.10. *Let M be an R-module. Then we have*

$$m_1 \wedge m_2 \wedge \cdots \wedge m_k = -m_{\tau_i(1)} \wedge m_{\tau_i(2)} \wedge \cdots \wedge m_{\tau_i(k)},$$

where $\tau_i = (i\ i+1) \in S_k$, for all $1 \leq i \leq k$ and all $m_j \in M$ for $1 \leq j \leq k$.

PROOF. The proof in the general case amounts to the following calculation in the case $k = 2$. For any $m, n \in M$, we have

$$0 = (m+n) \wedge (m+n) = m \wedge m + m \wedge n + n \wedge m + n \wedge n = m \wedge n + n \wedge m,$$

so $m \wedge n = -n \wedge m$. □

REMARK 9.4.11. The property that $m \wedge n = -n \wedge m$ for all $m, n \in M$ tells us directly that $m \wedge m = -m \wedge m$, and so $2m \wedge m = 0$, by taking $m = n$. In other words, if 2 is invertible in R, the submodule of $M \otimes_R M$ generated by tensors of the form $m \otimes n + n \otimes m$ contains the tensors of the form $m \otimes m$.

THEOREM 9.4.12. *Let M be a free R-module of rank n. Then the kth exterior power $\bigwedge^k M$ of M over R is a free R-module of rank $\binom{n}{k}$ for any $k \in \mathbb{Z}$, where we take $\binom{n}{k} = 0$ for $k > n$.*

PROOF. Let $m_1, \ldots, m_n$ be a basis of M. The 0th exterior power is just R, so the result holds for $k = 0$. For $k \geq 1$, we know that $M^{\otimes k}$ is R-free with a basis of elements of the form $m_{i_1} \otimes m_{i_2} \otimes \cdots \otimes m_{i_k}$ with $1 \leq i_j \leq n$ for each $1 \leq j \leq k$. Since we can switch the orders of the terms of elements of $\bigwedge^k M$ with only a change of sign, we have that $\bigwedge^k M$ is generated by the $m_{i_1} \wedge m_{i_2} \wedge \cdots \wedge m_{i_k}$ with $1 \leq i_1 \leq i_2 \leq \cdots \leq i_k \leq n$. But by definition of the exterior product, those elements with $i_j = i_{j+1}$ for some j are 0, so it is generated by those with $1 \leq i_1 < i_2 < \cdots < i_k \leq n$. The number of such elements is $\binom{n}{k}$.

It remains only to see R-linear independence. For this, fix $1 \leq i'_1 < i'_2 < \cdots < i'_k \leq n$, and define $f\colon M^k \to R$ as the unique R-multilinear map satisfying that $f(m_{i_1}, m_{i_2}, \ldots, m_{i_k})$ equals 0 unless $\{i_1, \ldots, i_k\} = \{i'_1, \ldots, i'_k\}$, in which case it is $\operatorname{sign}(\sigma)$ for $\sigma \in S_n$ such that $\sigma(i_j) = i'_j$ for $1 \leq j \leq k$ and σ fixes every other element of $\{1, 2, \ldots, n\}$. (Recall from Proposition 4.12.1 that the sign map can be defined independently of the definition of the determinant, so as to avoid circularity in our argument.) That this map is alternating can be easily checked: let $m = \sum_{i=1}^n r_i m_i \in M$, and consider

$$f(\ldots, m, m, \ldots) = \sum_{i=1}^n \sum_{j=1}^n r_i r_j f(\ldots, m_i, m_j, \ldots).$$

We then note that

$$f(\dots,m_i,m_j,\dots)+f(\dots,m_j,m_i,\dots)=0$$

for $i \neq j$ and $f(\dots,m_i,m_i,\dots)=0$ for all i to see that the sum is trivial. The map f then induces an element $F \in \mathrm{Hom}_R(\bigwedge^k M,R)$. Given some nontrivial R-linear combination x in $\bigwedge^k M$ of the generators $m_{r_1} \wedge m_{r_2} \wedge \cdots \wedge m_{r_k}$ with $r_1 < r_2 < \cdots < r_k$, the value $F(x)$ is also the coefficient of $m_{i_1} \wedge m_{i_2} \wedge \cdots \wedge m_{i_k}$ in the linear combination x. So, if $x=0$, then the linear combination must be the zero linear combination, which verifies R-linear independence. □

COROLLARY 9.4.13. *The R-module $\bigwedge^n R^n$ is one-dimensional with basis vector $e_1 \wedge e_2 \wedge \cdots \wedge e_n$, where $\{e_1,e_2,\dots,e_n\}$ is the standard basis of R^n.*

9.5. Graded rings

DEFINITION 9.5.1. A *graded ring A* is a ring determined by a sequence of abelian groups A_i for $i \geq 0$ and biadditive maps $\phi_{i,j}\colon A_i \times A_j \to A_{i+j}$ for $i,j \geq 0$ satisfying

$$\phi_{i+j,k}(\phi_{i,j}(r_i,r_j),r_k) = \phi_{i,j+k}(r_i,\phi_{j,k}(r_j,r_k))$$

for $r_i \in A_i$, $r_j \in A_j$, $r_k \in A_k$ and $i,j,k \geq 0$ and such that A_0 is a ring with multiplication $\phi_{0,0}$, where the additive group of A is $\bigoplus_{i=0}^{\infty} A_i$ and the multiplication on A is given by

$$\left(\sum_{i=0}^{\infty} r_i\right) \cdot \left(\sum_{i=0}^{\infty} s_i\right) = \sum_{k=0}^{\infty}\sum_{i=0}^{k} \phi_{i,k-i}(r_i,s_{k-i}),$$

where the sums are finite and $r_i, s_i \in A$ for all i. The group A_i is called the *degree i part*, or *ith graded piece*, of A, and an element of A_i is said to be *homogeneous* of degree i.

DEFINITION 9.5.2. A *graded algebra* over a commutative ring R is an R-algebra A that is a graded ring with structure map $R \to A_0 \cap Z(A)$.

DEFINITION 9.5.3. For a commutative ring R, an *homomorphism of graded R-algebras* $\psi\colon A \to B$ is a homomorphism of rings such that $\psi(A_i) \subseteq B_i$ for each $i \geq 0$.

Clearly if R has a grading, then R is a graded ring with respect to the resulting subgroups and maps.

DEFINITION 9.5.4. A *grading* on a ring R is a sequence of additive subgroups R_i with $i \geq 0$ such that R_0 is a subring and $R = \bigoplus_{i=0}^{\infty} R_i$ such that the multiplication on R restricts to maps $\phi_{i,j}\colon R_i \times R_j \to R_{i+j}$ for all $i,j \geq 0$. We say that R is *graded* by the R_i.

EXAMPLE 9.5.5. Any commutative or noncommutative polynomial ring R on a set X has a grading under which the nth graded piece is the R-span of of the words in X of length n. In fact, there are many possible gradings by assigning arbitrary choices of positive degrees to the different elements of X.

EXAMPLE 9.5.6. Given a ring A and an ideal I, we may form the graded ring $\mathrm{gr}_I A = \bigoplus_{n=0}^{\infty} I^n/I^{n+1}$, where the maps $I^i/I^{i+1} \times I^j/I^{j+1} \to I^{i+j}/I^{i+j+1}$ are given by $(x+I^{i+1}, y+I^{j+1}) \mapsto xy + I^{i+j+1}$. If A is an R-algebra, then $\mathrm{gr}_I A$ is a graded R-algebra via the map $R \to A/I$.

We can form an algebra out of the tensor powers of a module.

DEFINITION 9.5.7. For a commutative ring R and $n \geq 0$, the *nth tensor power* of an R-module M is $T^n(M) = M^{\otimes n} = M \otimes_R \cdots \otimes_R M$, the n-fold R-tensor product of M with itself, which is taken to be R if $n = 0$.

DEFINITION 9.5.8. For a commutative ring R and nonzero R-module M, the *tensor algebra* $T_R(M)$ of M is the graded R-algebra with ith graded piece $T^i(M)$ together with the unique R-bilinear maps $\phi_{i,j}\colon T^i(M) \times T^j(M) \to T^{i+j}(M)$ satisfying

$$\phi_{i,j}(m_1 \otimes \cdots \otimes m_i, n_1 \otimes \cdots n_j) = m_1 \otimes \cdots \otimes m_i \otimes n_1 \otimes \cdots n_j,$$

where the R-algebra structure map is the identity $R \to T^0(M)$

EXAMPLE 9.5.9. The R-tensor algebra of R is isomorphic to $R[x]$ as a graded R-algebra. That is, we have an isomorphism $\psi\colon R[x] \to T_R(R)$ of graded R-algebras uniquely determined by $\psi(x) = 1 \in T^1(R)$. More generally, the R-tensor algebra of R^n is isomorphic to $R\langle x_1, \ldots, x_n\rangle$ as a graded algebra (where the x_i have degree 1).

DEFINITION 9.5.10. A *graded ideal* of a graded ring is an ideal that has a homogeneous generating set.

The reader can verify the following.

LEMMA 9.5.11. *An ideal I of a graded ring A is homogeneous if and only if $I = \bigoplus_{n=0}^{\infty} I_n$, where $I_n = A_n \cap I$ for all $n \geq 0$.*

LEMMA 9.5.12. *The quotient of a graded R-algebra A by a homogeneous ideal I is a graded R-algebra with ith graded piece $A_i/(A_i \cap I)$, where A_i is the ith graded piece of A.*

DEFINITION 9.5.13. Let R be a commutative ring and M be an R-module.

a. The *symmetric algebra* $S_R(M)$ on a R-module M is the quotient of $T_R(M)$ by the homogeneous ideal generated by the elements $m \otimes n - n \otimes m$ with $m, n \in R$.

b. The nth graded piece $S^n(M)$ of $S_R(M)$ is called the nth *symmetric power* of M.

NOTATION 9.5.14. For M an R-module and $x, y \in T_R(M)$, the image of their product $x \otimes y$ in $S_R(M)$ is denoted $x \cdot y$.

EXAMPLE 9.5.15. The symmetric algebra $S_R(R^n)$ is isomorphic to $R[x_1, \ldots, x_n]$.

DEFINITION 9.5.16. Let R be a commutative ring and M be an R-module. The *exterior algebra* $\bigwedge_R M$ on M is the quotient of $T_R(M)$ by the homogeneous ideal generated by the elements $m \otimes m$ with $m \in R$.

NOTATION 9.5.17. For M an R-module and $x, y \in T_R(M)$, the image of their product $x \otimes y$ in $S_R(M)$ is denoted $x \wedge y$.

LEMMA 9.5.18. *The multiplication on $\bigwedge_R M$ for an R-module M satisfies $x \wedge x = 0$ and $x \wedge y = -y \wedge x$ for all $x, y \in \bigwedge_R M$.*

PROOF. For any $x, y \in T_R(M)$, we have

$$(x+y) \otimes (x+y) = x \otimes x + x \otimes y + y \otimes x + y \otimes y,$$

which reduces the problem to proving that $x \otimes x$ lies in the homogeneous ideal I generated by the $m \otimes m$ for $m \in R$. By the distributive property of multiplication, the result is further reduced to the case of simple tensors. For $m_1, \dots, m_n \in R$, we claim that

$$m_1 \otimes \cdots \otimes m_n \otimes m_1 \otimes \cdots \otimes m_n \in I,$$

and for this it suffices to show that

$$m_1 \otimes (m_2 \otimes \cdots \otimes m_n) \otimes m_1 \in I,$$

This is clear if $n = 1$. For $n \geq 2$, from the case $n = 1$ it follows that $m_1 \otimes m_2 - m_2 \otimes m_1 \in I$, which reduces us to showing that

$$m_1 \otimes (m_3 \otimes \cdots \otimes m_n) \otimes m_1 \in I,$$

which now follows by induction. □

The reader can now verify the following.

LEMMA 9.5.19. *For an R-module M and $n \geq 0$, the nth graded piece of $\bigwedge_R M$ is isomorphic to $\bigwedge^n M$ under the R-linear map that takes the image of $m_1 \otimes \cdots \otimes m_n$ to $m_1 \wedge \cdots \wedge m_n$ for $m_1, \dots, m_n \in M$.*

9.6. Determinants

In this section, R denotes a commutative ring.

DEFINITION 9.6.1. Let $n \geq 1$.

a. The *determinant* $\det(A)$ of a matrix $A \in M_n(R)$ with columns $v_1, \dots, v_n \in R^n$ is the unique element of R such that

$$v_1 \wedge v_2 \wedge \cdots \wedge v_n = \det(A) \cdot e_1 \wedge e_2 \wedge \cdots \wedge e_n,$$

where e_i denotes the ith element in the standard basis of R^n.

b. The *determinant map*

$$\det\colon M_n(R) \to R$$

is the map that takes a matrix to its determinant.

REMARK 9.6.2. The determinant map is an alternating, multilinear map if we view $M_n(R)$ as $\bigwedge^n R^n$ by taking a matrix to the wedge product $v_1 \wedge v_2 \wedge \cdots \wedge v_n$ of its columns $v_1, v_2, \dots, v_n$.

PROPOSITION 9.6.3. *The determinant map* $\det\colon M_n(R) \to R$ *satisfies*

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}$$

for any $A = (a_{ij}) \in M_n(R)$.

PROOF. Let $v_1, v_2, \dots, v_n$ denote the columns of $A = (a_{ij})$. Then $v_j = \sum_{i=1}^n a_{ij} e_i$. We have

$$v_1 \wedge v_2 \wedge \cdots \wedge v_n = \sum_{i_1=1}^n \sum_{i_2=1}^n \cdots \sum_{i_n=1}^n a_{i_1 1} a_{i_2 2} \cdots a_{i_n n} \cdot e_{i_1} \wedge e_{i_2} \wedge \cdots \wedge e_{i_n},$$

but note that all the terms such that the $j_1, j_2, \dots, j_n$ are not all distinct are zero. The remaining nonzero terms correspond to permutations $\sigma \in S_n$ with $\sigma(j) = i_j$ for each $1 \le j \le n$. We then have

$$\begin{aligned} v_1 \wedge v_2 \wedge \cdots \wedge v_n &= \sum_{\sigma \in S_n} a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} \cdot e_{\sigma(1)} \wedge e_{\sigma(2)} \wedge \cdots \wedge e_{\sigma(n)} \\ &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} \cdot e_1 \wedge e_2 \wedge \cdots \wedge e_n \\ &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \cdot e_1 \wedge e_2 \wedge \cdots \wedge e_n, \end{aligned}$$

the latter step coming from rearranging the terms and replacing σ by σ^{-1}. □

LEMMA 9.6.4. *Let $A, B \in M_n(R)$ for some $n \ge 1$. Then*

$$\det(AB) = \det(A)\det(B).$$

PROOF. Let v_i be the ith column of A, let w_i be the ith column of B, and let z_i be the ith column of AB. Then $Aw_i = z_i$ for all i. By Lemma 9.4.9, we then obtain

$$z_1 \wedge z_2 \wedge \cdots \wedge z_n = \det(A) w_1 \wedge w_2 \wedge \cdots \wedge w_n.$$

Since

$$w_1 \wedge w_2 \wedge \cdots \wedge w_n = \det(B) e_1 \wedge e_2 \wedge \cdots \wedge e_n,$$

the result holds. □

DEFINITION 9.6.5. Let R be a ring. Two matrices A and A' in $M_n(R)$ for some $n \ge 1$ are called *similar* if there exists a matrix $Q \in \mathrm{GL}_n(R)$ such that $A' = Q^{-1}AQ$.

REMARK 9.6.6. Let $T : R^n \to R^n$ be a linear transformation represented by the matrix A with respect to the standard basis of R^n. If $A' = Q^{-1}AQ$ for $Q \in \mathrm{GL}_n(R)$, then A' represents T with respect to the basis $B = \{v_1, \dots, v_n\}$ with $v_j = \sum_{i=1}^n q_{ij} e_i$ for $1 \le j \le n$. Conversely, any two matrices that each represent T with respect to some basis are similar.

Lemma 9.6.4 has the following corollary.

COROLLARY 9.6.7. *Let R be a commutative ring.*

a. For any $A \in \mathrm{GL}_n(R)$, we have $\det(A)\det(A^{-1}) = 1$.

b. Let A and B be similar matrices in $M_n(R)$. Then $\det(A) = \det(B)$.

LEMMA 9.6.8. *Let $A \in M_n(R)$, and let A^T denote its transpose. Then* $\det(A^T) = \det(A)$.

PROOF. Write $A = (a_{i,j})$. By Proposition 9.6.3, we have

$$\det(A^T) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n},$$

but $a_{\sigma(j)j} = a_{\sigma(j)\sigma^{-1}(\sigma(j))}$, so

$$a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} = a_{1\sigma^{-1}(1)} a_{2\sigma^{-1}(2)} \cdots a_{n\sigma^{-1}(n)}.$$

Noting that $\operatorname{sign}(\sigma) = \operatorname{sign}(\sigma^{-1})$ for all $\sigma \in S_n$, we have

$$\det(A^T) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma^{-1}) a_{1\sigma^{-1}(1)} a_{2\sigma^{-1}(2)} \cdots a_{n\sigma^{-1}(n)} = \det(A).$$

□

We also have the following standard properties of the determinant.

LEMMA 9.6.9. *Let $A \in M_n(R)$.*

a. Let B be a matrix obtained by switching either two rows or two columns of A. Then $\det(B) = -\det(A)$.

b. Let C be a matrix obtained by adding an R-multiple of one row (resp., column) of A to another row (resp., column). Then $\det(C) = \det(A)$.

c. Let D be a matrix obtained by multiplying one row or column of A by some $c \in R$. Then $\det(D) = c\det(A)$.

PROOF. By Lemma 9.6.8, it suffices to prove these for columns. Part a follows from the more general fact that

$$v_{\sigma(1)} \wedge v_{\sigma(2)} \wedge \cdots \wedge v_{\sigma(n)} = \operatorname{sign}(\sigma) v_1 \wedge v_2 \wedge \cdots \wedge v_n,$$

and part b follows from

$$\begin{aligned} &v_1 \wedge \cdots \wedge v_i \wedge \cdots \wedge (v_j + rv_i) \wedge \cdots \wedge v_n \\ &\qquad = (v_1 \wedge \cdots \wedge v_i \wedge \cdots \wedge v_j \wedge \cdots \wedge v_n) + r(v_1 \wedge \cdots \wedge v_i \wedge \cdots \wedge v_i \wedge \cdots \wedge v_n) \\ &\qquad = v_1 \wedge \cdots \wedge v_i \wedge \cdots \wedge v_j \wedge \cdots \wedge v_n. \end{aligned}$$

Part c follows from the multilinearity of the exterior product. □

LEMMA 9.6.10. *Let $A \in M_n(R)$ be a block diagonal matrix with $A \in M_{n_i}(R)$ for $1 \le i \le m$ and some $m \ge 1$. Then* $\det(A) = \prod_{i=1}^m \det(A_i)$.

PROOF. We have

$$Ae_1 \wedge Ae_2 \wedge \cdots \wedge Ae_n = \det(A_1)(e_1 \wedge \cdots \wedge e_{n_i}) \wedge \cdots \wedge \det(A_m)(e_{n-n_m+1} \wedge \cdots \wedge e_n),$$

as required. □

DEFINITION 9.6.11. For $A \in M_n(R)$ and $1 \le i, j \le n$, the (i,j)-minor of A is the matrix $A_{ij} \in M_{n-1}(R)$ obtained by removing the ith row and jth column from A. The (i,j)-cofactor of A is $(-1)^{i+j}\det(A_{ij})$.

PROPOSITION 9.6.12 (Cofactor expansion). *Let $A = (a_{ij}) \in M_n(R)$. Then for any i with $1 \le i \le n$, we have*

$$\det(A) = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det(A_{ij}),$$

and for any j with $1 \le j \le n$, we have

$$\det(A) = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det(A_{ij}).$$

PROOF. The first follows from the second by taking the transpose. So, fix j. Denote the ith column of A by v_i. Set

$$w_k^{(i)} = v_k - a_{ik}e_i$$

for each k, which is the column vector given by replacing the ith entry of v_k by zero. We may then view $w_1^{(i)}, \dots, w_{j-1}^{(i)}, w_j^{(i)}, \dots, w_n^{(i)}$ as the column vectors of the minor A_{ij} in the ordered basis $e_1, \dots, e_{i-1}, e_{i+1}, \dots, e_n$. In particular, we have

$$w_1^{(i)} \wedge \cdots \wedge w_{j-1}^{(i)} \wedge w_{j+1}^{(i)} \wedge \cdots \wedge w_n^{(i)} = \det(A)e_1 \wedge \cdots \wedge e_{i-1} \wedge e_{i+1} \wedge \cdots \wedge e_n.$$

We then have

$$\begin{aligned} v_1 \wedge v_2 \wedge \cdots \wedge v_n &= (-1)^{j-1} v_i \wedge v_1 \wedge \cdots \wedge v_{j-1} \wedge v_{j+1} \wedge \cdots \wedge v_n \\ &= (-1)^{j-1} \sum_{i=1}^n a_{ij} e_i \wedge v_1 \wedge \cdots \wedge v_{j-1} \wedge v_{j+1} \wedge \cdots \wedge v_n \\ &= (-1)^{j-1} \sum_{i=1}^n a_{ij} e_i \wedge w_1^{(i)} \wedge \cdots \wedge w_{j-1}^{(i)} \wedge w_{j+1}^{(i)} \wedge \cdots \wedge w_n^{(i)} \\ &= (-1)^{j-1} \sum_{i=1}^n a_{ij} \det(A_{ij}) e_i \wedge e_1 \wedge \cdots \wedge e_{i-1} \wedge e_{i+1} \wedge \cdots \wedge e_n \\ &= (-1)^{i+j} \sum_{i=1}^n a_{ij} \det(A_{ij}) e_1 \wedge e_2 \wedge \cdots \wedge e_n, \end{aligned}$$

where in the third equality we have applied Lemma 9.6.9(b). □

DEFINITION 9.6.13. Let $A \in M_n(R)$. The *adjoint matrix* to A is the matrix with (i, j)-entry $(-1)^{i+j} \det(A_{ji})$.

THEOREM 9.6.14. *Let $A \in M_n(R)$, and let B be its adjoint matrix. Then $AB = \det(A)I_n$.*

PROOF. The (i, j)-entry of AB is $\sum_{k=1}^n (-1)^{k+j} a_{ik} \det(A_{jk})$. If $i = j$, this is just $\det(A)$ by Proposition 9.6.12. If $i \neq j$, then the same proposition tells us that this equals the determinant of a matrix which has the ith row of the matrix obtained by replacing the jth row of A by the ith row of A. Since this matrix has two rows which are the same, its determinant is 0. □

COROLLARY 9.6.15. *A matrix $A \in M_n(R)$ is invertible if and only if $\det(A) \in R^\times$, in which case its inverse is $A^{-1} = \det(A)^{-1}B$, where B is the adjoint matrix to A.*

As any two similar matrices in $M_n(R)$ have the same determinant and any two matrices representing a linear transformation are similar, the following is well-defined.

DEFINITION 9.6.16. Let V be a free R-module of finite rank. The *determinant* of an R-module homomorphism $T : V \to V$ is the determinant of a matrix representing T with respect to an R-basis of V.

REMARK 9.6.17. Let $T : V \to V$ be a homomorphism of free R-modules, and let A be an R-algebra. Then we have an A-module homomorphism $\mathrm{id}_A \otimes T : A \otimes_F V \to A \otimes_F V$, which we usually denote more simply by T. It satisfies $T(a \otimes v) = a \otimes T(v)$ for any $a \in A$ and $v \in V$.

DEFINITION 9.6.18.

a. The *characteristic polynomial* of a matrix $A \in M_n(R)$ is $c_A(x) = \det(xI - A)$.

b. The *characteristic polynomial* of an R-module homomorphism $T\colon V \to V$ with V a free R-module of finite rank is $c_T(x) = \det(x\,\mathrm{id} - T)$, where id denotes the identity map on $F[x] \otimes_F V$.

DEFINITION 9.6.19. The *trace* of a matrix $A = (a_{ij}) \in M_n(R)$ is

$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii} \in R.$$

The trace is a homomorphism of additive groups.

LEMMA 9.6.20. *If* $A, B \in M_n(R)$*, then* $\operatorname{tr}(A+B) = \operatorname{tr}(A) + \operatorname{tr}(B)$.

LEMMA 9.6.21. *Let* $A \in M_n(R)$*. The constant coefficient of* $c_A(x)$ *is* $(-1)^n \det(A)$*, and the coefficient of* x^{n-1} *is* $-\operatorname{tr}(A)$.

PROOF. We have $c_A(0) = \det(-A) = (-1)^n \det(A)$. The second part is an easy consequence of the permutation formula for the determinant applied to $xI_n - A$, from which it is seen that only the term corresponding to the identity of S_n has degree at least $n-1$. This term is equal to $(x-a_{11})(x-a_{22})\cdots(x-a_{nn})$, and its x^{n-1}-coefficient is $-\operatorname{tr}(A)$. □

COROLLARY 9.6.22. *If A and B are similar matrices in* $M_n(R)$*, then* $\operatorname{tr}(A) = \operatorname{tr}(B)$.

The reader may also verify the following directly.

LEMMA 9.6.23. *Let* $A, B \in M_n(R)$*. Then* $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

9.7. Torsion and rank

DEFINITION 9.7.1. Let M be a module over an integral domain R. We say that $m \in M$ is an *R-torsion element* if there exists a nonzero element $a \in R$ with $am = 0$.

DEFINITION 9.7.2. Let M be a module over an integral domain R. Then M is said to be a *torsion module* if all of its nonzero elements are R-torsion elements.

LEMMA 9.7.3. *Let M be an module over an integral domain R. The set N of R-torsion elements of M is an R-submodule of M.*

PROOF. Let $n \geq 1$, and let $m_i \in M$ and $a_i, r_i \in R - \{0\}$ for $1 \leq i \leq n$ be such that $r_i m_i = 0$. Then $r = r_1 r_2 \cdots r_n$ is nonzero, and we have that $r \sum_{i=1}^{n} a_i m_i = 0$, so $\sum_{i=1}^{n} a_i m_i$ is R-torsion. □

DEFINITION 9.7.4. Let M be a module over an integral domain R. The *R-torsion submodule* M_{tor} of M is the set of R-torsion elements of M.

DEFINITION 9.7.5. Let R be a ring and M a left R-module. The *annihilator* of M in R is the left ideal

$$\operatorname{Ann}(M) = \{r \in R \mid rm = 0 \text{ for all } m \in M\}$$

of R.

The reader will easily verify the following.

LEMMA 9.7.6. *The annihilator* $\mathrm{Ann}(M)$ *of a left R-module over a ring R is a two-sided ideal of R.*

DEFINITION 9.7.7. Let R be a ring and M a left R-module. We say that an R-module M is *faithful* if $\mathrm{Ann}(M) = 0$.

REMARK 9.7.8. Let R be an integral domain and M an R-module. If $\mathrm{Ann}(M) \neq 0$, then M is R-torsion since any nonzero $r \in \mathrm{Ann}(M)$ satisfies $rm = 0$ for all $m \in M$.

LEMMA 9.7.9. *Let R be an integral domain, and let M be a finitely generated R-module. Then* $\mathrm{Ann}(M) \neq 0$ *if and only if M is R-torsion.*

PROOF. We may suppose that M is R-torsion. Let $m_1, \dots, m_n$ generate M, and let $r_1, \dots, r_n \in R - \{0\}$ be such that $r_i m_i = 0$ for $1 \leq i \leq n$. Then $r_1 r_2 \cdots r_n$ is a nonzero element of $\mathrm{Ann}(M)$. □

EXAMPLE 9.7.10. The abelian group $\bigoplus_{n=1}^{\infty} \mathbb{Z}/n\mathbb{Z}$ is both faithful and torsion as a $\mathbb{Z}$-module.

Let us introduce a general notion of rank for modules over integral domains.

DEFINITION 9.7.11. Let R be an integral domain, and let M be an R-module. The *rank* of M over R, or *R-rank* of M, is the largest nonnegative integer $n = \mathrm{rank}_R M$ such that M contains n elements that are linearly independent over R, if it exists. If $\mathrm{rank}_R M$ exists, then R is said to have *finite rank*, and otherwise it has *infinite rank*.

For free modules over integral domains, this agrees with the notion of rank defined above. We can give an alternative characterization of the rank. For this, we introduce the following lemma.

LEMMA 9.7.12. *Let* $\iota \colon M \to Q(R) \otimes_R M$ *be the R-module homomorphism defined by* $\iota(m) = 1 \otimes m$ *for* $m \in M$. *Then* $\ker \iota = M_{\mathrm{tor}}$.

PROOF. By Proposition 11.1.31, the module $Q(R) \otimes_R M$ is canonically isomorphic to the localization of M by $S = R - \{0\}$. The map ι becomes identified with the map $M \to S^{-1}M$ given by $m \mapsto \frac{m}{1}$. The definition of $S^{-1}M$ tells us that $\frac{m}{1} = 0$ if and only if there exists $r \in R - \{0\}$ such that $rm = 0$, which is to say $m \in M_{\mathrm{tor}}$. □

PROPOSITION 9.7.13. *Let R be an integral domain, and let M be an R-module. Then M has finite rank over R if and only if* $Q(R) \otimes_R M$ *is finite-dimensional over* $Q(R)$, *in which case*

$$\mathrm{rank}_R M = \dim_{Q(R)} Q(R) \otimes_R M.$$

PROOF. First, suppose that $m_1, m_2, \dots, m_n$ are n elements of M. First, suppose that the elements m_i are R-linearly dependent. Let $a_i \in R$ not all 0 be such that $\sum_{i=1}^{n} a_i m_i = 0$. Then

$$\sum_{i=1}^{n} a_i (1 \otimes m_i) = \sum_{i=1}^{n} a_i \otimes m_i = \sum_{i=1}^{n} 1 \otimes a_i m_i = 0,$$

so the elements $1 \otimes m_i$ are $Q(R)$-linearly dependent.

Conversely, suppose that the elements $1 \otimes m_i$ are $Q(R)$-linearly dependent. Let $\alpha_i \in Q(R)$ with $\sum_{i=1}^{n} \alpha_i \otimes m_i = 0$ and not all $\alpha_i = 0$. Let $d \in R$ be such that $a_i = d\alpha_i \in R$ for all i, and set $m = \sum_{i=1}^{n} a_i m_i$. We then have

$$1 \otimes m = \sum_{i=1}^{n} a_i \otimes m_i = d \sum_{i=1}^{n} \alpha_i \otimes m_i = 0,$$

so there exists $r \in R - \{0\}$ with $rm = 0$ by Lemma 9.7.12. That is, the m_i are R-linearly dependent. □

EXAMPLE 9.7.14. Set $R = \mathbb{Z}[x]$, and consider the ideal $I = (p, x)$, viewed as a left R-module. The usual method shows the existence of a map $I \otimes_R Q(R) \to Q(R)$ satisfying $f \otimes \frac{g}{h} \mapsto \frac{fg}{h}$. This map is clearly onto, so

$$\operatorname{rank}_R I = \dim_{Q(R)} I \otimes_R Q(R) \geq 1.$$

As $f \otimes \frac{g}{h} = x \otimes \frac{fg}{xh}$, the map $Q(R) \to I \otimes_R Q(R)$ given by $\frac{g}{h} \mapsto x \otimes \frac{g}{h}$ is onto, so $\operatorname{rank}_R I \leq 1$. Thus, I has R-rank 1, but it is not a free R-module as it cannot be generated by a single element.

9.8. Noetherian rings and modules

DEFINITION 9.8.1. Let R be a ring. A left R-module M is said to be *noetherian* if its set of submodules satisfies the ascending chain condition.

PROPOSITION 9.8.2. *A module M over a ring R is noetherian if and only if every submodule of M is finitely generated over R.*

PROOF. If every submodule of M is finitely generated, then the union of every ascending chain $\{N_i \mid i \geq 1\}$ of submodules of M is finitely generated, and each one of these generators is contained in some N_k, so they are all contained in the largest N_k among these. Thus, the union is actually equal to N_k, so the ACC holds.

On the other hand, if the ACC holds for M, then we can pick $m_1 \in M - \{0\}$ and then, if it exists, $m_{i+1} \in M$ with $m \notin M_i$ with $M_i = \sum_{j=1}^{i} Rm_i$ for each i. By definition, M_i is properly contained in M_{i+1}, so by the ACC, eventually we cannot continue the process, which is to say that for some k, we have $M_k = M$, or in other words that M is generated by $\{m_1, m_2, \ldots, m_k\}$. □

REMARK 9.8.3. Finitely generated modules need not be noetherian. A ring is left noetherian (i.e., satisfies the ascending chain condition on left ideals) if and only if it is noetherian as a left module over itself. Yet, any ring is finitely generated as a left module over itself, being that it is generated by 1.

LEMMA 9.8.4. *Let R be a ring, let M be an R-module, and let N be an R-submodule of M. If S is a generating set of N and T is a subset of M with image generating M/N, then $S \cap T$ generates M.*

PROOF. If $m \in M$, then there exist $n \in N$, $m_i \in T$, and $c_i \in R$ for $1 \leq i \leq j$ for some j such that

$$m = n + \sum_{i=1}^{j} c_i m_i,$$

and then there exist $n_i \in S$ and $c_i \in R$ for $1 \leq i \leq k$ for some k such that

$$n = \sum_{i=1}^{k} c_i n_i,$$

so n is an R-linear combination of the m_i and the n_i. That is, M is generated by $S \cup T$. □

COROLLARY 9.8.5. *Let R be a ring, let M be an R-module, and let N be a finitely generated R-submodule of M such that M/N is also finitely generated. Then M is finitely generated.*

LEMMA 9.8.6. *Let R be a ring and N be a submodule of an R-module M. Then M is noetherian if and only if both N and M/N are noetherian.*

PROOF. If M is noetherian, then N is noetherian by definition. Moreover, the inverse image P of any submodule O of M/N under the quotient map $\pi\colon M \to M/N$ is a submodule of M, hence generated by some finite set S. Then $\pi(S)$ generates O, so we conclude that M/N is noetherian.

If N and M/N are both noetherian and P is a submodule of M, then $P \cap N$ and $P/(P \cap N)$ are finitely generated as submodules of N and M/N, respectively, so P is finitely generated by Corollary 9.8.5. That is, M is noetherian. □

COROLLARY 9.8.7. *Finite direct sums of noetherian modules are noetherian.*

PROOF. If $M = N \oplus N'$ for R-modules M, N, and N', then $N' = M/N$, so by the lemma, M is noetherian if N and N' are. The result then follows by induction on the number of summands. □

PROPOSITION 9.8.8. *Every finitely generated left module over a left noetherian ring is noetherian.*

PROOF. Let M be a finitely generated left module over a noetherian ring R, and let N be a submodule of M. Since M is finitely generated, there is a surjective R-module homomorphism $R^n \to M$ for some n. Let P be the inverse image of N in R^n. The module N is generated by the image of any set of generators of P under the quotient map $P \to N$. So, we need only show that any submodule P of R^n is finitely generated, which is to say that R^n is left noetherian. This is true as R is a left noetherian R-module, and R^n is the direct sum of n copies of R. □

THEOREM 9.8.9 (Hilbert's basis theorem). *The polynomial ring $R[x]$ over a commutative noetherian ring R is noetherian.*

PROOF. Let I be an ideal of $R[x]$. We must show that I is finitely generated. Let L be the set the leading coefficients of the elements of I. Then L is clearly an ideal: if $a \in L$ is the leading coefficient of $f \in I$ and $r \in R$, then ra is the leading coefficient of $rf \in I$, and if $a, b \in L$ are the leading coefficients of f and g, respectively, then $x^{\deg g} f + x^{\deg f} g \in I$ has leading coefficient $a + b$. Since R is noetherian, there exist $a_1, a_2, \ldots, a_k \in J$ such that $R = (a_1, a_2, \ldots, a_k)$. Let $f_i \in I$ of degree $n_i \geq 0$ have leading coefficient a_i for $1 \leq i \leq k$. Let $n = \max\{n_i \mid 1 \leq i \leq k\}$.

Next, for $m \geq 0$, let L_m be the set of all leading coefficients of polynomials in I of degree m. This, again, is clearly an ideal of R, so we have $L_m = (b_{m,1}, b_{m,2}, \ldots, b_{m,l_m})$ for some $l_m \geq 1$ and $b_{m,i} \in J_m$ for $1 \leq i \leq l_m$. For each such i, let $g_{m,i} \in I$ be a polynomial of degree m with leading coefficient $b_{m,i}$. We claim that I is generated by

$$X = \{f_i \mid 1 \le i \le k\} \cup \bigcup_{m=0}^{n} \{g_{m,i} \mid 1 \le i \le l_m\}.$$

Let J be the ideal of $R[x]$ generated by X, which is contained in I. Let $h \in I$, and let $c \in L$ be its leading coefficient. We want to show that $h \in J$. Write $c = \sum_{i=1}^{k} r_i a_i$ with $r_i \in R$. If $d \ge n$, then c is the leading coefficient of

$$h' = \sum_{i=1}^{k} r_i x^{d-n_i} f_i \in J,$$

so $h - h'$ has degree less than d. We can then replace h by $h - h'$ and repeat the process until $d < n$.

We are reduced to showing that if $h \in I$ has degree $d < n$ and leading coefficient c, then $h \in J$. In this case, we have $c = \sum_{i=1}^{l_d} s_i b_{d,i}$ with $s_i \in S$, and c is the leading coefficient of

$$h' = \sum_{i=1}^{l_d} s_i g_{d,i} \in J$$

Then $h - h' \in J$ has degree less than d. Replacing $h - h'$ by h and repeating the process, we see that $h \in J$. □

COROLLARY 9.8.10. *Let R be a noetherian ring. Then $R[x_1, x_2, \ldots, x_n]$ is noetherian for every $n \ge 1$.*

PROPOSITION 9.8.11. *Any finitely generated (commutative) algebra over a field is noetherian.*

PROOF. Let $\mathscr{A}$ be a finitely generated algebra over a field F. If $a_1, a_2, \ldots, a_n$ generate $\mathscr{A}$ as an F-algebra, then we have a surjective F-algebra homomorphism

$$\pi \colon F[x_1, x_2, \ldots, x_n] \to \mathscr{A}$$

defined by $\pi(f) = f(a_1, a_2, \ldots, a_n)$. Thus $\mathscr{A}$ is a quotient of $\mathscr{B} = F[x_1, x_2, \ldots, x_n]$. If I is an ideal of $\mathscr{A}$, then $\pi^{-1}(I)$ is an ideal of $\mathscr{B}$. As $\mathscr{B}$ is noetherian, this ideal is finitely generated, and the images of its generators generate I. □

We next consider modules that satisfy the descending chain condition.

DEFINITION 9.8.12. Let X be a set with a partial ordering $\le$. A *descending chain* on X is an ascending chain with respect to the opposite partial ordering $\ge$ defined by $x \ge y$ if and only if $y \le x$ for $x, y \in X$.. We say that X satisfies the descending chain condition, or *DCC*, if it satisfies the ACC with respect to $\ge$.

DEFINITION 9.8.13. We say that a module over a ring R is *artinian* if its set of submodules satisfies the descending chain condition.

EXAMPLE 9.8.14. Any finite-dimension vector spaces over a field F is an artinian F-module.

LEMMA 9.8.15. *Let R be a ring and N be a submodule of an R-module M. Then M is artinian if and only if both N and M/N are artinian.*

PROOF. That M being artinian implies N and M/N are artinian is straightforward. If N and M/N are artinian and $(M_i)_{i\geq 1}$ is a descending chain in M, then there exists $k \geq 0$ such that $M_i \cap N = M_k \cap N$ and $(M_i+N)/N = (M_k+N)/N$ for all $i \geq k$. But this can only happen if $M_i = M_k$ for all $i \geq k$ as well: if $m \in M_k$, then $m \in M_i + N$, so $m = m' + n$ for some $m' \in M_i$ and $n \in N$, but then $n = m - m' \in M_k \cap N \subseteq M_i$, and therefore $m = m' + n \in M_i$. □

LEMMA 9.8.16. *If $\mathfrak{m}$ is a maximal ideal in a noetherian ring, then $R/\mathfrak{m}^n$ is an artinian R-module.*

PROOF. Note that $R/\mathfrak{m}$ is a field, hence artinian as an R-module. By induction on $n \geq 1$, we may suppose that $R/\mathfrak{m}^{n-1}$ is artinian as an R-module. By Lemma 9.8.15, it suffices to show that $\mathfrak{m}^{n-1}/\mathfrak{m}^n$ is Artinian over R. Since R is noetherian, $\mathfrak{m}^{n-1}$ is a finitely generated R-module, and the images in $\mathfrak{m}^{n-1}/\mathfrak{m}^n$ of any list of generators span it as an $R/\mathfrak{m}$-vector space. Since it is artinian as an $R/\mathfrak{m}$-module, it is also artinian as an R-module. □

9.9. Modules over PIDs

LEMMA 9.9.1. *Let R be a PID. Any finitely generated R-submodule of $Q(R)$ is cyclic.*

PROOF. Let M be an R-module generated by some subset $\{\alpha_1, \dots, \alpha_n\}$ of $Q(R)$. Let $d \in R - \{0\}$ be such that $d\alpha_i \in R$ for all i. Then $d \colon M \to dM$ is an isomorphism, and dM is an ideal of R, hence principal. That is, dM is cyclic as an R-module, so M is as well. □

PROPOSITION 9.9.2. *Let R be a PID. Let V be an n-dimensional $Q(R)$-vector space, and let M be a finitely generated R-submodule of V. Then there exists a basis $\{v_1, v_2, \dots, v_n\}$ of V and $k \leq n$ such that M is a free R-module with R-basis $\{v_1, v_2, \dots, v_k\}$.*

PROOF. We suppose without loss of generality that M is nonzero. Pick a nonzero element $m_1 \in M$. Recall that R is noetherian as it is a PID. Then $Q(R)m_1$ is a 1-dimensional $Q(R)$-vector space, and M is noetherian being that it is R-finitely generated, so $M \cap Q(R)m_1$ is R-finitely generated. Since $Q(R)m_1$ is a 1-dimensional $Q(R)$-vector space with R-submodule $M \cap Q(R)m_1$, Lemma 9.9.1 tells us that $M \cap Q(R)m_1 = Rv_1$ for some $v_1 \in Q(R)m_1$. Set $\bar{M} = M/Rv_1$. This is an R-submodule of the $(n-1)$-dimensional vector space $\bar{V} = V/Q(R)v_1$, since if $x \in M$ is such that $x + Rv_1$ is in the kernel of $\bar{M} \to \bar{V}$, then there exists $\alpha \in Q(R)$ such that $x = \alpha v_1$. But then $x \in M \cap Q(R)m_1$, which is to say $x \in Rv_1$.

Now, by induction on n, there exist $v_2, \dots, v_k \in M$ for some $k \geq 1$ such that $v_2 + Rv_1, \dots, v_k + Rv_1$ form an R-basis of $\bar{M}$. Then $v_1, v_2, \dots, v_k \in M$ generate M by Lemma 9.8.4, and we claim they are R-linearly independent. That is, if $\sum_{i=1}^k c_i v_i = 0$ for some $c_i \in R$, then

$$\sum_{i=2}^{k} c_i(v_i + Rm_1) = 0,$$

and so $c_i = 0$ for $2 \leq i \leq k$. As $v_1 \neq 0$, this forces $c_1 = 0$ as well. To finish, we merely extend $\{v_1, \dots, v_k\}$ to a $Q(R)$-basis $\{v_1, \dots, v_n\}$ of V, noting that an R-linearly independent subset of V is also $Q(R)$-linearly independent. □

COROLLARY 9.9.3. *Every finitely generated, torsion-free module over a principal ideal domain is free.*

PROOF. Let M be a finitely generated, torsion-free module over a PID R. We have seen in Lemma 9.7.12 that the canonical map $M \to Q(R) \otimes_R M$ is injective, in that $M_{\mathrm{tor}} = 0$. The result is then immediate from Proposition 9.9.2, as $Q(R) \otimes_R M$ is a finite-dimensional $Q(R)$-vector space. □

COROLLARY 9.9.4. *Any submodule of free module of rank n over a principal domain is free of rank at most n.*

PROOF. Let R be a PID. Let M be a free R-module of rank n, and let N be an R-submodule of M. Then $N \to Q(R) \otimes_R M$ is injective, so we can apply Proposition 9.9.2. □

PROPOSITION 9.9.5. *Let M be a finitely generated module over a principal ideal domain R. Then $M \cong R^r \oplus M_{\mathrm{tor}}$, where r is the rank of M.*

PROOF. Note that M/M_{tor} is free of finite rank by Corollary 9.9.3, so isomorphic to R^r for some r. By Proposition 5.7.26, we have that that $M \cong M_{\mathrm{tor}} \oplus R^r$. □

LEMMA 9.9.6. *Let R be a principal ideal domain, let $\pi \in R$ be an irreducible element. Let $k \geq 1$ and set $\bar{R} = R/(\pi^k)$. Then any free $\bar{R}$-submodule F of a finitely generated $\bar{R}$-module M is a direct summand of M. If F is maximal, then $M \cong F \oplus C$, where $\pi^{k-1}C = 0$.*

PROOF. We work by induction on k. Let M be an $\bar{R}$-module and F a free submodule of M. If $k = 1$, then M is a finite-dimensional $\bar{R}$-vector space. Then F is a direct summand of M, since any basis of it extends to a basis of M.

Now take $k \geq 2$. Suppose first that A is a maximal free $\bar{R}$-submodule of M. Consider the subgroup

$$N = \{m \in M \mid \pi^{k-1}m = 0\}.$$

We have $\pi A \subseteq N$, which is an $R/(\pi^{k-1})$-module. By induction on k, the free $R/(\pi^{k-1})$-module πA is a direct summand of N. We have $N = \pi A \oplus C$ for some $\bar{R}$-submodule C of A.

Any set of representatives in M of an $R/(\pi)$-basis of M/N is $\bar{R}/(\pi^k)$-linearly independent, hence a basis of a free $\bar{R}$-submodule of M. The map $A/\pi A \to M/N$ is injective. If it were not surjective, we could by Lemma 11.2.14 extend the image of a basis of A to a basis of M/N and lift to obtain a free $\bar{R}$-module of higher rank containing F. Therefore, it is an isomorphism. In particular, $A + N = M$, so $A + C = M$. Note also that $A \cap N = \pi A$, so $A \cap C = 0$. Thus $M = A \oplus C$.

It remains to show that an arbitrary free $\bar{R}$-module F is a direct summand of M. It suffices to show that F is a direct summand of a maximal free $\bar{R}$-submodule A. For this, note that the map $F/\pi F \to A/\pi A$ is injective, since if $a \in \pi A \cap F$, then $\pi^{k-1}a = 0$, so $a \in \pi F$ by the freeness of F. We may then choose a set X that is a basis of a complement of $F/\pi F$ in $A/\pi A$. Again by Lemma 11.2.14, any lift of X to a linearly independent subset of A will span an $\bar{R}$-complement to F. Thus F is a direct summand of A. □

We are now ready to prove the structure theorem for finitely generated modules over principal ideal domains.

THEOREM 9.9.7 (Structure theorem for finitely generated modules over PIDs). *Let R be a PID, and let M be a finitely generated R-module.*

a. There exist unique nonnegative integers r and k and nonzero proper principal ideals $I_1 \subseteq I_2 \subseteq \cdots \subseteq I_k$ of R such that

$$M \cong R^r \oplus R/I_1 \oplus R/I_2 \oplus \cdots \oplus R/I_k.$$

b. There exist unique nonnegative integers r and l, and for $1 \le i \le l$, distinct nonzero prime ideals $\mathfrak{p}_i$ of R and positive integers $v_{i,1} \ge v_{i,2} \ge \cdots \ge v_{i,m_i}$ for some $m_i \ge 1$ such that

$$M \cong R^r \oplus \bigoplus_{i=1}^{l} \bigoplus_{j=1}^{m_i} R/\mathfrak{p}_i^{v_{i,j}}. \tag{9.9.1}$$

Moreover, r and l are unique, and the tuple $(\mathfrak{p}_i, (v_{i,j})_j)_i$ is unique up to ordering in i.

PROOF. By Proposition 9.9.5, it suffices to consider the case that M is torsion. Note that the uniqueness of r in parts (a) and (b) follows from the fact that $r = \dim_{Q(R)} Q(R) \otimes_R M = \operatorname{rank}_R M$. So, let M be a finitely generated torsion R-module.

We first demonstrate the existence of a decomposition as in part b. Let $c \in R$ be a generator of the annihilator $\operatorname{Ann}(M)$ of M. Since we have unique factorization in R, we may write $c = u\pi_1^{k_1} \cdots \pi_l^{k_l}$ with $u \in R^\times$ and with $\pi_1, \dots, \pi_r$ distinct irreducible elements of M and $k_1, \dots, k_;$ positive integers for some $k \ge 0$. By the Chinese remainder theorem, we have an isomorphism

$$R/(c) \cong \prod_{i=1}^{l} R/(\pi_i^{k_i})$$

of rings, which in turn provides a direct sum decomposition

$$M = M/cM \cong M \otimes_R R/(c) \cong \bigoplus_{i=1}^{l} M \otimes_R R/(\pi_i^{k_i}) \cong \bigoplus_{i=1}^{l} M/\pi_i^{k_i} M.$$

In particular, we are reduced to the case that M is a module over the local ring $R/(\pi^k)$ for some irreducible element π of R.

For the moment, suppose that M is a nonzero finitely generated $R/(\pi^k)$-module for some $k \ge 1$. Note that if $k = 1$, then M is simply a finite dimensional $R/(\pi)$-vector space, so a choice of basis gives a direct sum decomposition $M \cong \bigoplus_{i=1}^m R/(\pi)$ for some m. For general k, let F be a maximal free $R/(\pi^k)$-submodule of M. By Lemma 9.9.6, we have $M = F \oplus C$, where C is a finitely generated $R/(\pi^{k-1})$-module. By induction on k, this gives the decomposition of part b.

We next prove the existence in part a using the decomposition in part b. Let us take π_i to be an irreducible element generating $\mathfrak{p}_i$ for each i. For $j \ge 1$, set $b_j = \pi_1^{v_{1,j}} \pi_2^{v_{2,j}} \cdots \pi_l^{v_{l,j}}$. By construction, we have that $b_{j+1} \mid b_j$ for each $j \ge 1$. Set $I_j = (b_j)$, and let k be maximal such that $I_k \ne R$ or zero if all $I_k = R$. Applying the Chinese remainder theorem again to see that

$$\bigoplus_{i=1}^{l} R/\pi_i^{v_{i,j}} \cong R/I_j$$

we obtain a decomposition as in part a.

Next, we exhibit uniqueness. If $\mathfrak{p} = (\pi)$ is a nonzero prime ideal of R and $v \geq 0$, then the map $R/\mathfrak{p} \to \mathfrak{p}^v/\mathfrak{p}^{v+1}$ induced by multiplication by π^v is an isomorphism. Let $N = R/\mathfrak{q}^w$ for a nonzero prime ideal $\mathfrak{q}$ and $w \geq 0$. Note that $\mathfrak{p}^v + \mathfrak{q}^w = R$ if $\mathfrak{p} \neq \mathfrak{q}$, being that $\mathfrak{p}$ and $\mathfrak{q}$ are generated by coprime elements. We therefore have

$$\mathfrak{p}^v N/\mathfrak{p}^{v+1}N \cong \mathfrak{p}^v/((\mathfrak{p}^{v+1}+\mathfrak{q}^w)\cap\mathfrak{p}^v) \cong \begin{cases} R/\mathfrak{p} & \text{if } \mathfrak{p} = \mathfrak{q} \text{ and } v < w \\ 0 & \text{if } \mathfrak{p} \neq \mathfrak{q} \text{ or } v \geq w. \end{cases}$$

For any $v \geq 1$, we then have that

$$\mathfrak{p}_i^v M/\mathfrak{p}_i^{v+1} M \cong (R/\mathfrak{p}_i)^u,$$

where u is the number of j such that $v < v_{i,j}$. Thus, the $\mathfrak{p}_i$ and $v_{i,j}$ in any decomposition of M as in part b are the same.

Finally, we reduce the uniqueness in part a to the known uniqueness of part b. Given any decomposition $M \cong R/I_1 \oplus \cdots \oplus R/I_k$ as in part a, we can again obtain a decomposition of M into R modulo power of prime ideals, applying CRT to expand out each R/I_j. Since $I_j \supseteq I_{j+1}$, this decomposition then satisfies $I_j = \mathfrak{p}_1^{v_{1,j}}\mathfrak{p}_2^{v_{2,j}}\cdots\mathfrak{p}_l^{v_{l,j}}$ with $v_{i,j} \geq v_{i,j+1}$ for each $j \geq 0$. Thus, the decomposition is as in part b, and by its uniqueness, we obtain the uniqueness of the decomposition in part a. □

REMARK 9.9.8. The structure theorem for finitely generated abelian groups is the special case of the structure theorem for finitely generated modules over a PID for the PID $\mathbb{Z}$.

DEFINITION 9.9.9. Let R be a PID, and let M be a finitely generated R-module.

a. The ideals $I_1, I_2, \ldots, I_k$ associated to M by Theorem 9.9.7a are called the *invariant factors* of M.

b. The prime powers $\mathfrak{p}_i^{v_{i,j}}$ associated to M by Theorem 9.9.7b are called the *elementary divisors* of M.

9.10. Canonical forms

DEFINITION 9.10.1. Let V be a vector space over a field F, and let $T\colon V \to V$ be an F-linear transformation.

a. An *eigenvector* v of T with *eigenvalue* $\lambda \in F$ is an element $v \in V - \{0\}$ such that $Tv = \lambda v$.

b. An element $\lambda \in F$ is called an *eigenvalue* of T if there exists an eigenvector in V with eigenvalue λ.

c. The *eigenspace* of F for $\lambda \in F$ of T is the nonzero subspace

$$E_\lambda(T) = \{v \in V \mid T(v) = \lambda v\}$$

of V.

Note that $E_\lambda(T) = \ker(T - \lambda\,\mathrm{id}_V)$.

LEMMA 9.10.2. *Let V be a vector space over a field F. The following are equivalent for an F-linear transformation $T\colon V \to V$ and $\lambda \in F$:*

i. $E_\lambda(T) \neq 0$,

ii. λ *is an eigenvalue of* T, *and*

iii. $c_T(\lambda) = 0$.

PROOF. The first two are equivalent by definition. Moreover, $T - \lambda \operatorname{id}_V$ has a nonzero kernel if and only if $c_T(\lambda) = \det(\lambda \operatorname{id}_V - T) = 0$. □

TERMINOLOGY 9.10.3. We may speak of eigenvectors, eigenvalues, and eigenspaces $E_\lambda(A)$ of a matrix in $A \in M_n(F)$, taking them to be the corresponding objects for the linear transformation $T\colon F^n \to F^n$ that A represents.

The following is the key to the application of the structure theorem for modules over PIDs to linear algebra.

NOTATION 9.10.4. If $T\colon V \to V$ is a linear transformation and $f = \sum_{i=1}^k c_i x^i \in F[x]$, then we set

$$f(T) = \sum_{i=1}^n c_i T^i\colon V \to V,$$

where $T^i\colon V \to V$ denotes the i-fold composition of T with itself.

REMARK 9.10.5. If T is represented by a matrix A and $f = \sum_{i=1}^k c_i x^i \in F[x]$, then $f(T)$ is represented by

$$f(A) = \sum_{i=1}^k c_i A^i \in M_n(F).$$

DEFINITION 9.10.6. Let $T\colon V \to V$ be an F-linear endomorphism of an F-vector space V. The $F[x]$-module structure endowed on V by T is that which satisfies $f(x) \cdot v = f(T)v$ for all $T \in F[x]$.

This construction gives us one way to define the minimal polynomial of a linear transformation.

DEFINITION 9.10.7.

a. Let V be a finite-dimensional F-vector space. The *minimal polynomial* $m_T(x)$ of a linear transformation $T\colon V \to V$ is the unique monic generator of the annihilator $\operatorname{Ann}(V)$ under the $F[x]$-module structure on V induced by T.

b. The minimal polynomial $m_A(x)$ is the minimal polynomial of the linear transformation $T\colon F^n \to F^n$ that A represents with respect to the standard basis of F^n.

LEMMA 9.10.8. *The minimal polynomial of an endomorphism T of a finite-dimensional vector space V divides the characteristic polynomial of T.*

PROOF. Let $v \in V$. Then $\det(xI - T) \cdot v = \det(T - T) \cdot v = 0$ by definition, so $c_T(x) \in \operatorname{Ann}(V)$. □

LEMMA 9.10.9. *If A and B are similar matrices in $M_n(F)$, then $c_A(x) = c_B(x)$ and $m_A(x) = m_B(x)$.*

PROOF. Suppose $Q \in \mathrm{GL}_n(F)$ is such that $B = QAQ^{-1}$. Then $xI - B = Q(xI - A)Q^{-1}$, so

$$c_B(x) = \det(xI - B) = \det(xI - A) = c_A(x).$$

Moreover, if $g \in F[x]$ is such that $g(A)v = 0$ for all $v \in F^n$, then

$$g(B)Qv = Qg(A)v = 0$$

for all $v \in F^n$, so $g(B)$ annihilates F^n as well. By symmetry, we have $m_A(x) = m_B(x)$. □

LEMMA 9.10.10. *Let A be a block diagonal matrix with blocks $A_i \in M_{n_i}(F)$ for $1 \le i \le m$ and some $m \ge 1$. Then*

$$c_A(x) = \prod_{i=1}^{m} c_{A_i}(x),$$

while $m_A(x)$ is the least common multiple of the $m_{A_i}(x)$ with $1 \le i \le m$.

Suppose that we endow a finite-dimensional F-vector space V with the structure of an $F[x]$-module through a linear transformation $T\colon V \to V$. Since $F[x]$ is a PID, the structure theorem for modules over a PID tells us that there exists an $F[x]$-module isomorphism

$$V \cong \bigoplus_{i=1}^{m} F[x]/(f_i),$$

where $m \ge 0$ and the $f_i \in F[x]$ are monic, nonconstant polynomials for $1 \le i \le m$ such that $f_{i+1} \mid f_i$ for $1 \le i < m$.

LEMMA 9.10.11. *Let $f = \sum_{i=0}^{n} c_i x^i \in F[x]$ be a monic polynomial of degree $n \ge 1$. With respect to the ordered basis $\{1, x, \dots, x^{n-1}\}$ of $V = F[x]/(f)$ as an F-vector space, the linear transformation given by multiplication by x on V is represented by the matrix*

$$A_f = \begin{pmatrix} 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & & 0 & -c_1 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & -c_{n-2} \\ 0 & \cdots & 0 & 1 & -c_{n-1} \end{pmatrix}.$$

(This matrix is taken to be $(-c_0)$ if $n = 1$.)

PROOF. Let $T_f\colon V \to V$ be the linear transformation given by left multiplication by x. Note that $T_f(x^i) = x^{i+1}$ for $0 \le i \le n-2$ and

$$T_f(x^{n-1}) = x^n = -\sum_{i=0}^{n-1} c_i x^i.$$

Thus, if $A_f = (a_{i,j})$, we have $a_{i+1,i} = 1$ for $1 \le i \le n-1$ and $a_{i,n} = -c_{i-1}$ for $1 \le i \le n$, and all other entries are zero. □

DEFINITION 9.10.12. For any monic, nonconstant $f \in F[x]$, the matrix A_f of Lemma 9.10.11 is known as the *companion matrix* to f.

LEMMA 9.10.13. *If $f \in F[x]$ is nonconstant and monic, then $c_{A_f}(x) = m_{A_f}(x) = f$.*

PROOF. The case $n = 1$ is clear. Write $f = xg + c_0$ for some $g \in F[x]$. By induction of the degree of f, we have

$$c_{A_f} = \det(xI - A_f) = x\det(xI - A_g) + (-1)^{n-1}c_0\det(-I_{n-1}) = xg + c_0 = f.$$

□

We can make the following definition as a consequence of the structure theorem for $F[x]$-modules.

DEFINITION 9.10.14. Let V be a finite-dimensional F-vector space and $T\colon V \to V$ an F-linear transformation. Write $V \cong \bigoplus_{i=1}^m F[x]/(f_i)$ for some $m \geq 0$ and monic $f_i \in F[x]$ with $f_i \mid f_{i+1}$ for all $i < m$. The *rational canonical form* of T is the block-diagonal matrix

$$\begin{pmatrix} A_{f_1} & & & \\ & A_{f_2} & & \\ & & \ddots & \\ & & & A_{f_m} \end{pmatrix},$$

where A_{f_i} is the companion matrix of f_i.

REMARK 9.10.15. The rational canonical form represents T with respect to the basis of V determined by taking the image under the isomorphism $\bigoplus_{i=1}^m F[x]/(f_i) \xrightarrow{\sim} V$ of the ordered basis of the direct sum given by concatenating the bases $\{1, x, \ldots, x^{\deg(f_i)-1}\}$ of the ith summands in order of increasing i.

We also note the following.

REMARK 9.10.16. By definition of rational canonical form, a matrix in rational canonical form in one field is already in rational canonical form in any extension field.

DEFINITION 9.10.17. The *rational canonical form* of a matrix $A \in M_n(F)$ is the rational canonical form of the linear transformation that A represents with respect to the standard basis of F^n.

REMARK 9.10.18. By Remark 9.10.15 and the change of basis theorem, A is similar to its rational canonical form. Moreover, two matrices are similar if and only if they have the same rational canonical form, since similar n-by-n matrices give rise to isomorphic $F[x]$-module structures on F^n and conversely.

DEFINITION 9.10.19. The *invariant factors* of $A \in M_n(F)$ are the invariant factors of F^n viewed as an $F[x]$-module via the linear transformation represented by A with respect to the standard basis.

As a simple consequence of Lemmas 9.10.13 and 9.10.10, we have the following.

LEMMA 9.10.20. *Let $f_1, f_2, \ldots, f_m$ be the invariant factors of a matrix A (with $f_i \mid f_{i+1}$ for $i < m$). Then $m_A(x) = f_m(x)$ and $c_A(x) = \prod_{i=1}^m f_i(x)$.*

As a consequence, the irreducible divisors of $c_A(x)$ and $m_A(x)$ are the same. In particular, we have:

COROLLARY 9.10.21. *Let $A \in M_n(F)$ and $\lambda \in F$. The following are equivalent:*

i. An element $\lambda \in F$ is an eigenvalue of $A \in M_n(F)$.

ii. The polynomial $x - \lambda$ divides the minimal polynomial $m_A(x)$.

iii. The polynomial $x - \lambda$ divides the characteristic polynomial $c_A(x)$.

This lemma is at times enough to calculate the rational canonical form of a matrix.

EXAMPLES 9.10.22. Let $A \in M_n(F)$.

a. If $f = c_A(x)$ is a product of distinct monic, irreducible polynomials, then the rational canonical form of A is A_f.

b. If $f = m_A(x)$ has degree n, then the rational canonical form of A is A_f.

c. If $f = c_A(x) = m_A(x)^d$ and $g = m_A(x)$ is irreducible of degree $\frac{n}{d}$, then A has d invariant factors of the form A_g.

d. Note that

$$\begin{pmatrix} 0 & & & \\ & 0 & & \\ & & 0 & 0 \\ & & 1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 0 & & \\ 1 & 0 & & \\ & & 0 & 0 \\ & & 1 & 0 \end{pmatrix}$$

are both 4-by-4 matrices in rational canonical form with characteristic polynomial x^4 and minimal polynomial x^2.

Recall that we have a second decomposition of V for the $F[x]$-module structure given by T. That is, there exist distinct monic, irreducible polynomials $p_1(x), p_2(x), \dots, p_l(x)$ and positive integers $v_{i,j}$ for $1 \le j \le m_i$ for some $m_i \ge 1$ for $1 \le i \le l$ such that

$$V \cong \bigoplus_{i=1}^{l} \bigoplus_{j=1}^{m_i} F[x]/(p_i(x)^{v_{i,j}}).$$

If the field F contains all the roots of $c_A(x)$, then it contains all the roots of the p_i, so being irreducible, these polynomials must be linear. This occurs, for instance, if F is algebraically closed. Let us assume this is the case and write $p_i(x) = x - \lambda_i$ for some $\lambda_i \in F$.

LEMMA 9.10.23. *Let $V = F[x]/((x-\lambda)^n)$ for some $\lambda \in F$ and $n \ge 0$. The linear transformation given by multiplication by x on V is represented by the matrix*

$$J_{\lambda,n} = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}$$

with respect to the ordered basis $\{(x-\lambda)^{n-1}, (x-\lambda)^{n-2}, \cdots, x-\lambda, 1\}$ of V.

PROOF. The linear transformation $T\colon V \to V$ that is multiplication by x satisfies

$$T((x-\lambda)^j) = (x-\lambda)^{j+1} + \lambda(x-\lambda)^j$$

for all j, with $(x-\lambda)^j = 0$ in V for $j \geq n$. The result follows. □

DEFINITION 9.10.24. A matrix $J_{\lambda,n}$ of the form Lemma 9.10.23 is called a Jordan block of dimension n for λ.

DEFINITION 9.10.25. Let V be a finite-dimensional F-vector space, and let $T\colon V \to V$ an F-linear transformation such that $c_T(x)$ splits in F. Write $V \cong \bigoplus_{i=1}^m F[x]/((x-\lambda_i)^{n_i})$ for some $\lambda_i \in F$ and $n_i \geq 1$ for $1 \leq i \leq m$ and some $m \geq 1$. The *Jordan canonical form of T* is a block-diagonal matrix

$$\begin{pmatrix} J_{\lambda_1,n_1} & & & \\ & J_{\lambda_2,n_2} & & \\ & & \ddots & \\ & & & J_{\lambda_m,n_m} \end{pmatrix}$$

whereJ_{λ_i,n_i} is the Jordan block of dimension n_i for λ_i.

TERMINOLOGY 9.10.26. If the characteristic polynomial of $c_T(x)$ splits in F, we say that T has a Jordan canonical form over F.

The Jordan canonical form is unique up to ordering of the Jordan blocks.

DEFINITION 9.10.27. The *Jordan canonical form* of a matrix $A \in M_n(F)$ is the Jordan canonical form of the linear transformation that A represents with respect to the standard basis of F^n.

REMARK 9.10.28. Every (square) matrix has a rational canonical form, while every matrix over an algebraically closed field has a Jordan canonical form.

PROPOSITION 9.10.29. *Suppose that $T\colon V \to V$ has a Jordan canonical form over F. Then $\lambda \in F$ is an eigenvalue of T if and only if it is a diagonal entry of the Jordan canonical form of T.*

PROOF. Consider the isomorphism $V \cong \bigoplus_{i=1}^m F[x]/((x-\lambda_i)^{n_i})$ The image $v \in V$ of $(x-\lambda_i)^{n_{i-1}}$ in the ith term of the right-hand side of the above isomorphism is an eigenvector with eigenvalue λ_i. On the other hand, if $\lambda \neq \lambda_i$, then $(x-\lambda)f \neq 0$ for nonzero $f \in F[x]/((x-\lambda_i)^{n_i})$, so λ is not an eigenvalue of T. □

EXAMPLES 9.10.30. Let $A \in M_n(F)$, and suppose that the characteristic polynomial of A splits in F.

a. If $c_A(x)$ is a product of distinct linear factors, then the Jordan canonical form of A is diagonal with entries the distinct eigenvalues of A.

b. If $m_A(x)$ is a product of distinct linear factors, then the Jordan canonical form of A is diagonal with entries that are all distinct eigenvalues of A.

c. If $m_A(x)$ has degree n, then the rational canonical form of A is block-diagonal with Jordan blocks J_{λ_i,n_i}, where the λ_i are all distinct.

d. If $c_A(x) = (x-\lambda)^n$ for some $\lambda \in F$ and $m_A(x) = x - \lambda$, then $A = J_{\lambda,n}$.

DEFINITION 9.10.31. Suppose that $T\colon V \to V$. The generalized eigenspace of $\lambda \in F$ for T is the $F[x]$-submodule

$$\{v \in V \mid (T-\lambda)^n(v) = 0 \text{ for some } n \geq 0\}$$

of V.

REMARK 9.10.32. The generalized eigenspace of $\lambda \in F$ under $T\colon V \to V$ contains the eigenspace $E_\lambda(T)$ of λ.

EXAMPLE 9.10.33. The generalized eigenspace of

$$A = \begin{pmatrix} J_{\lambda_1,n_1} & & & \\ & J_{\lambda_2,n_2} & & \\ & & \ddots & \\ & & & J_{\lambda_m,n_m} \end{pmatrix} \in M_n(F)$$

is the span of the elements e_i of the standard basis of F^n for which the ith diagonal entry of A is λ_i.

The following is an easy consequence of the example just given.

PROPOSITION 9.10.34. *Let $T\colon V \to V$ be a linear transformation. Then V is the direct sum of its nontrivial generalized eigenspaces if and only if $c_T(x)$ splits in F.*

We provide one example of how to obtain the rational and Jordan canonical forms of a matrix.

EXAMPLE 9.10.35. Let

$$A = \begin{pmatrix} 2 & -2 & 8 \\ 0 & 3 & -5 \\ 0 & 0 & 2 \end{pmatrix} \in M_3(\mathbb{Q}).$$

We have

$$c_A(x) = \det \begin{pmatrix} x-2 & 2 & -8 \\ 0 & x-3 & 5 \\ 0 & 0 & x-2 \end{pmatrix} = (x-2)^2(x-3).$$

So, $\mathbb{Q}^3$ is the direct sum of its generalized eigenspaces for 2 and 3. We compute that

$$A-2 = \begin{pmatrix} 0 & -2 & 8 \\ 0 & 1 & -5 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad (A-2)^2 = \begin{pmatrix} 0 & -2 & 10 \\ 0 & 1 & -5 \\ 0 & 0 & 0 \end{pmatrix},$$

so $\ker(A-2) = \langle e_1 \rangle$ and $\ker((A-2)^2) = \langle e_1, 5e_2 + e_3 \rangle$, where we use angle brackets to denote the $\mathbb{Q}$-span. Note that $(A-2)(5e_2+e_3) = -2e_1$. Similarly, we compute $\ker(A-3) = \langle 2e_1 - e_2 \rangle$. The Jordan canonical form of the matrix A is then

$$\begin{pmatrix} 2 & 1 & \\ & 2 & \\ & & 3 \end{pmatrix}$$

with respect to the basis $\langle -2e_1, 5e_2+e_3, 2e_1-e_2\rangle$.

By the Chinese remainder theorem, we have $F^3 \cong F[x]/((x-2)^2(x-3))$ under the $F[x]$-module structure on F^3 induced by A. Note that

$$(x-2)^2(x-3) = (x^2-4x+4)(x-3) = x^3-7x^2+16x-12.$$

To find a basis of the rational canonical form

$$\begin{pmatrix} 0 & 0 & 12 \\ 1 & 0 & -16 \\ 0 & 1 & 7 \end{pmatrix}$$

of A, we pick a vector v that generates F^3 as an $F[x]$-module, and then A is in rational canonical form with respect to the basis $\{v, Av, A^2v\}$. To find v, note that

$$(A-2)(A-3) = \begin{pmatrix} 0 & -2 & 8 \\ 0 & 1 & -5 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} -1 & -2 & 8 \\ 0 & 0 & -5 \\ 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

so $v = e_3$ works, and one possible basis is $\{e_3, 8e_1-5e_2+e_3, 42e_1-25e_2+4e_3\}$.

DEFINITION 9.10.36. We say that a matrix $A \in F^n$ is *diagonalizable* if it is similar to a diagonal matrix. A linear transformation $T\colon V \to V$ is diagonalizable if and only if T is representable by a diagonal matrix with respect to some basis of V.

Clearly, a linear transformation $T\colon V \to V$ is diagonalizable if and only if V is the direct sum of its distinct eigenspaces. The following is then a special case of Proposition 9.10.34.

PROPOSITION 9.10.37. *A linear transformation $T\colon V \to V$ is diagonalizable if and only if $m_T(x)$ splits in F.*

CHAPTER 10

Topics in Galois theory

10.1. Norm and trace

DEFINITION 10.1.1. Let E/F be a finite extension of fields. For $\alpha \in E$, let $m_\alpha \colon E \to E$ denote the F-linear transformation defined by left multiplication by α.

a. The norm map $\mathrm{N}_{E/F} \colon E \to F$ is defined by $\mathrm{N}_{E/F}(\alpha) = \det m_\alpha$ for $\alpha \in E$.

b. The trace map $\mathrm{Tr}_{E/F} \colon E \to F$ is defined by $\mathrm{Tr}_{E/F}(\alpha) = \mathrm{tr}\, m_\alpha$ for $\alpha \in E$.

REMARK 10.1.2. For a finite field extension E/F, the trace map $\mathrm{Tr}_{E/F}$ is a homomorphism, and the norm map $\mathrm{N}_{E/F}$ is a homomorphism to $F^\times$ upon restriction to $E^\times$.

PROPOSITION 10.1.3. *Let E/F be a finite extension of fields, and let $\alpha \in E$. Let $f \in F[x]$ be the minimal polynomial of α over F, let $d = [F(\alpha) : F]$, let $s = [E : F(\alpha)]$, and let $\overline{F}$ be an algebraic closure of F. Suppose that f factors in $\overline{F}[x]$ as*

$$f = \prod_{i=1}^{d} (x - \alpha_i)$$

for some $\alpha_1, \dots, \alpha_d \in \overline{F}$. Then the characteristic polynomial of m_α is f^s, and we have

$$\mathrm{N}_{E/F}(\alpha) = \prod_{i=1}^{d} \alpha_i^s \quad \text{and} \quad \mathrm{Tr}_{E/F}(\alpha) = s \sum_{i=1}^{d} \alpha_i.$$

PROOF. If $\{\beta_1, \dots, \beta_s\}$ is a basis for $E/F(\alpha)$, then $\{\beta_i \alpha^j \mid 1 \le i \le s, 0 \le j \le d-1\}$ is a basis for E/F. The matrix A representing m_α with respect to this basis (with the lexicographical ordering on the pairs (i,j)) is block diagonal with s blocks all equal to the matrix for $m_\alpha \colon F(\alpha) \to F(\alpha)$ for the ordered basis $\{1, \alpha, \dots, \alpha^{d-1}\}$. As we have an isomorphism of fields $F(\alpha) \cong F[x]/(f)$ fixing F under which α is sent to the coset of x, the latter matrix is the companion matrix A_f.

By Lemma 9.10.10, we have $\mathrm{char}\, m_\alpha = f^s$. Lemma 9.6.21 tells us that

$$f^s = x^d - \mathrm{tr}(m_\alpha) x^{d-1} + \cdots + (-1)^d \det(m_\alpha),$$

By expanding out the factorization of f^s in $\overline{F}[x]$, we see that $\mathrm{N}_{E/F}\, \alpha$ and $\mathrm{Tr}_{E/F}\, \alpha$ are as stated in this case. □

We can also express the norm as a power of a product of conjugates and the trace as a multiple of a sum of conjugates.

PROPOSITION 10.1.4. *Let E/F be a finite field extension, and let $t = [E : F]_{\mathrm{i}}$ be its degree of inseparability. Then, for $\alpha \in E$, we have*

$$\mathrm{N}_{E/F}(\alpha) = \prod_{\sigma \in \mathrm{Emb}_F(E)} \sigma\alpha^t \quad \text{and} \quad \mathrm{Tr}_{E/F}(\alpha) = t \sum_{\sigma \in \mathrm{Emb}_F(E)} \sigma\alpha.$$

PROOF. The distinct conjugates of α in a fixed algebraic closure $\overline{F}$ of F are exactly the $\tau\alpha$ for $\tau \in \mathrm{Emb}_F(F(\alpha))$. These $\tau\alpha$ are the distinct roots of the minimal polynomial of α over F, each occuring with multiplicity the degree $[F(\alpha) : F]_{\mathrm{i}}$ of insparability of $F(\alpha)/F$. Now, as in the proof of Lemma 6.10.23, each of these embeddings extends to $[E : F(\alpha)]_{\mathrm{s}}$ distinct embeddings of E into $\overline{F}$, and each extension $\sigma \in \mathrm{Emb}_F(E)$ of τ sends α to $\tau(\alpha)$. By Proposition 10.1.3, we have

$$\mathrm{N}_{E/F}\,\alpha = \prod_{\tau \in \mathrm{Emb}_F(F(\alpha))} (\tau\alpha)^{[E:F(\alpha)][F(\alpha):F]_{\mathrm{i}}} = \prod_{\sigma \in \mathrm{Emb}_F(E)} \sigma\alpha^{[E:F(\alpha)]_{\mathrm{i}}[F(\alpha):F]_{\mathrm{i}}} = \prod_{\sigma \in \mathrm{Emb}_F(E)} \sigma\alpha^t,$$

and similarly for the trace. □

We have the following immediate corollary.

COROLLARY 10.1.5. *Let E/F be a finite separable extension of fields. Then, for $\alpha \in E$, we have*

$$\mathrm{N}_{E/F}(\alpha) = \prod_{\sigma \in \mathrm{Emb}_F(E)} \sigma\alpha \quad \text{and} \quad \mathrm{Tr}_{E/F}(\alpha) = \sum_{\sigma \in \mathrm{Emb}_F(E)} \sigma\alpha.$$

We also have the following.

PROPOSITION 10.1.6. *Let K/F be a finite field extension and E be an intermediate field in the extension. Then we have*

$$\mathrm{N}_{K/F} = \mathrm{N}_{E/F} \circ \mathrm{N}_{K/E} \quad \text{and} \quad \mathrm{Tr}_{K/F} = \mathrm{Tr}_{E/F} \circ \mathrm{Tr}_{K/E}.$$

PROOF. Since $[K : F]_{\mathrm{i}} = [K : E]_{\mathrm{i}}[E : F]_{\mathrm{i}}$, it suffices by Proposition 10.1.4 to show that

$$\prod_{\delta \in \mathrm{Emb}_F(K)} \delta\alpha = \prod_{\sigma \in \mathrm{Emb}_F(E)} \sigma\left(\prod_{\tau \in \mathrm{Emb}_E(K)} \tau\alpha\right).$$

We extend each σ to an automorphism $\tilde{\sigma}$ of $\overline{F}$ fixing F. We then have

$$\prod_{\sigma \in \mathrm{Emb}_F(E)} \sigma\left(\prod_{\tau \in \mathrm{Emb}_E(K)} \tau\alpha\right) = \prod_{\sigma \in \mathrm{Emb}_F(E)} \prod_{\tau \in \mathrm{Emb}_E(K)} (\tilde{\sigma} \circ \tau)\alpha. \tag{10.1.1}$$

For the trace map, we simply replace the products by sums.

Let $\sigma, \sigma' \in \mathrm{Emb}_F(E)$ and $\tau, \tau' \in \mathrm{Emb}_E(K)$, and suppose that

$$\tilde{\sigma} \circ \tau = \tilde{\sigma}' \circ \tau' \in \mathrm{Emb}_F(K) \tag{10.1.2}$$

Since $\tilde{\sigma} \circ \tau|_F = \sigma|_F$, we have that $\sigma = \sigma'$. Since $\tilde{\sigma}$ is an automorphism, we then apply its inverse to (10.1.2) to obtain $\tau = \tau'$. As there are

$$|\mathrm{Emb}_F(E)||\mathrm{Emb}_E(K)| = [E : F]_s[K : E]_s = [K : F]_s = |\mathrm{Emb}_F(K)|$$

terms of the product in (10.1.1), we have the result. □

EXAMPLE 10.1.7. The norm for the extension $\mathbb{Q}(\sqrt{d})/\mathbb{Q}$, where d is a square-free integer, is given by

$$\mathrm{N}_{\mathbb{Q}(\sqrt{d})/\mathbb{Q}}(x+y\sqrt{d}) = (x+y\sqrt{d})(x-y\sqrt{d}) = x^2-dy^2$$

for $x,y \in \mathbb{Q}$.

EXAMPLE 10.1.8. For $a,b,c \in \mathbb{Q}$, we have

$$\begin{aligned}\mathrm{N}_{\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}}&(a+b\sqrt[3]{2}+c(\sqrt[3]{2})^2)\\ &= (a+b\sqrt[3]{2}+c(\sqrt[3]{2})^2)(a+b\omega\sqrt[3]{2}+c\omega^2(\sqrt[3]{2})^2)(a+b\omega^2\sqrt[3]{2}+c\omega(\sqrt[3]{2})^2)\\ &= a^3+2b^3+4c^3-6abc,\end{aligned}$$

for ω a primitive cube root of unity. The trace is simpler:

$$\mathrm{Tr}_{\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}}(a+b\sqrt[3]{2}+c(\sqrt[3]{2})^2) = 3a.$$

DEFINITION 10.1.9. A E-valued *linear character of a group G* is a group homomorphism $\chi\colon G \to E^\times$, where E is a field.

DEFINITION 10.1.10. We say that a set of E-valued linear characters X of a group G is *E-linearly independent* if it linearly independent as a subset of the E-vector space of functions $G \to E$.

THEOREM 10.1.11. *Any set of E-valued linear characters $G \to E^\times$ of a group G is E-linearly independent.*

PROOF. Let X be a set of linear characters $G \to E^\times$. Suppose by way of contradiction that $m \geq 1$ is minimal such that there m distinct, linearly dependent elements of G. Choose $a_i \in E$ and $\chi_i \in X$ with $1 \leq i \leq m$ for which

$$\sum_{i=1}^{m} a_i\chi_i = 0.$$

Also, let $h \in G$ be such that $\chi_1(h) \neq \chi_m(h)$. Set $b_i = a_i(\chi_i(h)-\chi_m(h))$ for $1 \leq i \leq m-1$. For any $g \in G$, we then have

$$\sum_{i=1}^{m-1} b_i\chi_i(g) = \sum_{i=1}^{m} a_i(\chi_i(h)-\chi_m(h))\chi_i(g)\sum_{i=1}^{m} a_i\chi_i(hg) - \chi_m(h)\sum_{i=1}^{m} a_i\chi_i(g) = 0.$$

Since $b_1 \neq 0$ and $\sum_{i=1}^{m-1} b_i\chi_i$ has only $m-1$ terms, this contradicts the existence of m. □

In the case of cyclic extensions, the kernels of the norm map bears a simple description.

THEOREM 10.1.12 (Hilbert's Theorem 90). *Let E/F be a finite cyclic extension of fields, and let σ be a generator of its Galois group. Then*

$$\ker \mathrm{N}_{E/F} = \left\{ \frac{\sigma(\beta)}{\beta} \mid \beta \in E^\times \right\}.$$

PROOF. Set $n = [E : F]$. Let $\beta \in E$, and note that

$$\mathrm{N}_{E/F}\left(\frac{\sigma(\beta)}{\beta}\right) = \prod_{i=0}^{n-1} \frac{\sigma^{i+1}(\beta)}{\sigma^i(\beta)} = \frac{\mathrm{N}_{E/F}(\beta)}{\mathrm{N}_{E/F}(\beta)} = 1.$$

Next, suppose that $\alpha \in \ker \mathrm{N}_{E/F}$, and set

$$x_\gamma = \gamma + \alpha\sigma(\gamma) + \alpha\sigma(\alpha)\sigma^2(\gamma) + \cdots + \alpha\sigma(\alpha)\cdots\sigma^{n-2}(\alpha)\sigma^{n-1}(\gamma)$$

for $\gamma \in E$. The elements of $\mathrm{Gal}(E/F)$, which is to say the powers of σ, are distinct E-valued characters on $E^\times$, and therefore they are E-linearly independent. Thus, there exists $\gamma \in E^\times$ such that $x_\gamma \neq 0$. We then note that

$$\alpha\sigma(x_\gamma) = \alpha\sigma(\gamma) + \alpha\sigma(\alpha)\sigma^2(\gamma) + \cdots + \alpha\sigma(\alpha)\cdots\sigma^{n-2}(\alpha)\sigma^{n-1}(\gamma) + \mathrm{N}_{E/F}(\alpha)\gamma = x_\gamma,$$

so $\alpha = \frac{\sigma(x_\gamma^{-1})}{x_\gamma^{-1}}$, finishing the proof. □

There is also an additive form of Hilbert's Theorem 90, which describes the kernel of the trace. We leave the proof to the reader.

PROPOSITION 10.1.13 (Additive Hilbert's Theorem 90). *Let E/F be a finite cyclic extension of fields, and let σ be a generator of its Galois group. Then*

$$\ker \mathrm{Tr}_{E/F} = \{\sigma(\beta) - \beta \mid \beta \in E\}.$$

10.2. Discriminants

In this section, we give a second treatment of discriminants.

DEFINITION 10.2.1. Let F be a field and V a finite-dimensional F-vector space. A *F-bilinear form* is a F-bilinear map $\psi\colon V \times V \to F$.

DEFINITION 10.2.2. A F-bilinear form ψ on a F-vector space V is said to be *symmetric* if $\psi(v,w) = \psi(w,v)$ for all $v,w \in V$.

EXAMPLE 10.2.3. Given a matrix $Q \in M_n(F)$, we can define a bilinear form on F^n by

$$\psi(v,w) = v^T Q w$$

for $v,w \in F^n$ which is symmetric if and only if Q is.

EXAMPLE 10.2.4. If E/F is a finite extension of fields, then $\psi\colon E \times E \to F$ defined by

$$\psi(\alpha,\beta) = \mathrm{Tr}_{E/F}(\alpha\beta)$$

for $\alpha,\beta \in E$ is a symmetric F-bilinear form on E.

DEFINITION 10.2.5. The *discriminant* $\mathrm{D}(\psi)$ of a bilinear form ψ on a finite dimensional F-vector space V relative to an ordered basis $(v_1,\ldots,v_n)$ of V is the determinant of the matrix with (i,j) entry equal to $\psi(v_i,v_j)$.

LEMMA 10.2.6. *Let $\psi\colon V \times V \to F$ be a F-bilinear form on a finite-dimensional vector space V of dimension $n \geq 1$. Let $v_1, \ldots, v_n \in V$, and let $T\colon V \to V$ be a F-linear transformation. Then*

$$\det(\psi(Tv_i, Tv_j)) = (\det T)^2 \cdot \det(\psi(v_i, v_j))$$

PROOF. It suffices to show this in the case that the v_i form a basis of V, since in this case, for any $w_1, \ldots, w_n \in V$ there exists a linear transformation $U\colon V \to V$ with $U(v_i) = w_i$ for all i. We then have

$$\det(\psi(Tw_i, Tw_j)) = \det(\psi(TUv_i, TUv_j)) = \det(TU)^2 \det(\psi(v_i, v_j)) = \det(T)^2 \det(\psi(w_i, w_j)).$$

So, assume that the v_i form a basis of V, and let $A = (a_{ij})$ denote the matrix of T with respect to the ordered basis $(v_1, \ldots, v_n)$ of V. We have $Tv_i = \sum_{k=1}^n a_{ki} v_k$ for each i, and therefore we have

$$\psi(Tv_i, Tv_j) = \sum_{k=1}^n a_{ki} \sum_{l=1}^n a_{lj} \psi(v_k, v_l).$$

As matrices, we then have

$$(\psi(Tv_i, Tv_j)) = A^T(\psi(v_i, v_j))A,$$

and the result follows as $\det T = \det A = \det A^T$. □

REMARK 10.2.7. It follows from Lemma 10.2.6 that the discriminant of a bilinear form with respect to a basis is independent of its ordering, since a permutation matrix has determinant ±1.

DEFINITION 10.2.8. Let E/F be a finite extension of fields. The *discriminant* $\mathrm{D}(\beta_1, \ldots, \beta_n)$ of E/F relative to an ordered basis $(\beta_1, \ldots, \beta_n)$ of E as a F-vector space is the discriminant of the bilinear form

$$(\alpha, \beta) \mapsto \mathrm{Tr}_{E/F}(\alpha\beta)$$

relative to the basis.

PROPOSITION 10.2.9. *Let E/F be a finite separable extension of fields. Then the discriminant of E/F relative to an ordered basis $(\beta_1, \ldots, \beta_n)$ of E satisfies*

$$\mathrm{D}(\beta_1, \ldots, \beta_n) = (\det(\sigma_i \beta_j))^2,$$

where $\{\sigma_1, \ldots, \sigma_n\}$ is the set of embeddings of E in an algebraic closure of F that fix F.

PROOF. Note that

$$\mathrm{Tr}_{E/F}(\beta_i \beta_j) = \sum_{i=1}^n \sigma_k(\beta_i)\sigma_k(\beta_j),$$

so the matrix $(\mathrm{Tr}_{E/F}(\beta_i\beta_j))$ equals $Q^T Q$, where $Q \in M_n(E)$ satisfies $Q_{ij} = \sigma_i(\beta_j)$. □

DEFINITION 10.2.10. Let F be a field, and let $\alpha_1, \ldots, \alpha_n \in F$. The *Vandermonde matrix* for $\alpha_1, \ldots, \alpha_n$ is

$$Q(\alpha_1, \ldots, \alpha_n) = \begin{pmatrix} 1 & \alpha_1 & \cdots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \cdots & \alpha_2^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & \alpha_n & \cdots & \alpha_n^{n-1} \end{pmatrix}.$$

LEMMA 10.2.11. *Let F be a field, and let $Q(\alpha_1,\dots,\alpha_n)$ be the Vandermonde matrix for elements $\alpha_1,\dots,\alpha_n$ of F. Then*

$$\det Q(\alpha_1,\dots,\alpha_n) = \prod_{1\le i<j\le n} (\alpha_j-\alpha_i).$$

PROOF. We work by induction on $n \ge 1$, the case $n = 1$ asserting the obvious fact that $\det Q(\alpha) = 1$ for any $\alpha \in F$. To compute the determinant of $Q = Q(\alpha_1,\dots,\alpha_n)$, in order of descending $i \le n-1$ subtract α_1 times the ith column of Q from the $(i+1)$th column, which leaves the determinant unchanged. We then obtain

$$\begin{aligned}\det Q &= \begin{vmatrix} 1 & 0 & \cdots & 0 \\ 1 & \alpha_2-\alpha_1 & \cdots & \alpha_2^{n-2}(\alpha_2-\alpha_1) \\ \vdots & \vdots & & \vdots \\ 1 & \alpha_n-\alpha_1 & \cdots & \alpha_n^{n-2}(\alpha_n-\alpha_1)\end{vmatrix} = \begin{vmatrix} \alpha_2-\alpha_1 & \cdots & \alpha_2^{n-2}(\alpha_2-\alpha_1) \\ \vdots & & \vdots \\ \alpha_n-\alpha_1 & \cdots & \alpha_n^{n-2}(\alpha_n-\alpha_1)\end{vmatrix} \\ &= \prod_{i=2}^{n}(\alpha_i-\alpha_1)\cdot Q(\alpha_2,\dots,\alpha_n),\end{aligned}$$

and the result now follows by induction. □

PROPOSITION 10.2.12. *Suppose that E/F is a separable extension of degree n, and let $\alpha \in E$ be such that $E = F(\alpha)$. Then*

$$\mathrm{D}(1,\alpha,\dots,\alpha^{n-1}) = \mathrm{D}(f),$$

where $f \in F[x]$ is the minimal polynomial of α.

PROOF. Let $\sigma_1,\sigma_2,\dots,\sigma_n$ be the embeddings of E in a fixed algebraic closure of F, and set $\alpha_i = \sigma_i(\alpha)$. Then $\sigma_i(\alpha^{j-1}) = \alpha_i^{j-1}$, so Lemmas 10.2.9 and 10.2.11 tell us that

$$\mathrm{D}(1,\alpha,\dots,\alpha^{n-1}) = \det Q(\alpha_1,\alpha_2,\dots,\alpha_n)^2 = \prod_{1\le i<j\le n}(\alpha_j-\alpha_i)^2.$$

The latter term is just $\mathrm{D}(f)$. □

DEFINITION 10.2.13. Let F be a field and $f = \sum_{i=0}^n a_i x^i \in F[x]$. The *derivative* $f' \in F[x]$ of f is $f' = \sum_{i=1}^n i a_i x^{i-1}$.

REMARK 10.2.14. An irreducible polynomial $f \in F[x]$ is inseparable if and only if $f' = 0$.

PROPOSITION 10.2.15. *Suppose that E/F is a separable extension of degree n, and let $\alpha \in E$ be such that $E = F(\alpha)$, and let $f \in F[x]$ be the minimal polynomial of α. Then*

$$\mathrm{D}(f) = (-1)^{\frac{n(n-1)}{2}} N_{E/F}(f'(\alpha)),$$

where $f' \in F[x]$ is the derivative of f.

PROOF. Let $\alpha_1,\dots,\alpha_n$ be the conjugates of α in an algebraic closure $\overline{F}$ of F. Then

$$f'(x) = \sum_{i=1}^{n}\prod_{\substack{j=1\\ j\neq i}}^{n}(x-\alpha_j),$$

so we have

$$f'(\alpha_i) = \prod_{\substack{j=1 \\ j\neq i}}^{n} (\alpha_i - \alpha_j)$$

for each i, and the conjugates of $f'(\alpha)$ in $\overline{F}$ are the $f'(\alpha_i)$. We then have

$$N_{E/F}(f'(\alpha)) = \prod_{i=1}^{n} \prod_{\substack{j=1 \\ j\neq i}}^{n} (\alpha_i - \alpha_j) = (-1)^{\frac{n(n-1)}{2}} \mathrm{D}(f)$$

□

COROLLARY 10.2.16. *Let L/F be a finite separable extension of fields. Then the discriminant of L/F relative to an ordered basis $(\beta_1, \beta_2, \dots, \beta_n)$ of L is nonzero.*

PROOF. Since L/F is separable, there exists $\alpha \in L$ such that $L = F(\alpha)$. Then $(1, \alpha, \dots, \alpha^{n-1})$ is an ordered basis of L/F, and there exists an invertible F-linear transformation $T\colon L \to L$ with $T(\alpha^{i-1}) = \beta_i$ for $1 \le i \le n$. By Lemma 10.2.6, we have that

$$\mathrm{D}(\beta_1, \beta_2, \dots, \beta_n) = (\det T)^2 \mathrm{D}(1, \alpha, \dots, \alpha^{n-1}).$$

It follows Proposition 10.2.12 that $\mathrm{D}(1, \alpha, \dots, \alpha^{n-1}) \neq 0$, so we have the result. □

REMARK 10.2.17. Together, Lemma 10.2.6 and Corollary 10.2.16 tell us that the discriminant of a finite separable field extension L/F (relative to an ordered basis) reduces to a element of $F^\times/F^{\times 2}$ that is independent of the choice of basis.

10.3. Extensions by radicals

DEFINITION 10.3.1. Let F be a field. A *Kummer extension E* of F is one that is given by adjoining roots of elements of F.

NOTATION 10.3.2. For a field F of characteristic not dividing $n \ge 1$, we let μ_n denote the group of nth roots of unity in a given algebraic closure of F.

PROPOSITION 10.3.3. *Let F be a field, and let $\overline{F}$ be a fixed algebraic closure of F. Let $n \ge 1$ and $a \in F$. Let $E = F(\alpha)$ for $\alpha \in \overline{F}$ with $\alpha^n = a$, and let $d \ge 1$ be minimal such that $\alpha^d \in F$.*

a. The extension E/F is Galois if and only if $\mathrm{char}\, F$ *does not divide d and E contains μ_d.*

b. If E/F is Galois and $\mu_d \subset F$, then the map

$$\chi_a \colon \mathrm{Gal}(E/F) \xrightarrow{\sim} \mu_d, \qquad \chi_a(\sigma) = \frac{\sigma(\alpha)}{\alpha}.$$

is an isomorphism of groups.

PROOF. The minimal polynomial f of α divides $x^d - \alpha^d$ but not $x^m - \alpha^m$ for any m dividing d. If μ_d has order d, then f is separable as $x^d - \alpha^d$ is. If μ_d has order m for some m properly dividing d, then f divides $x^d - \alpha^d = (x^m - \alpha^m)^{d/m}$ but not $x^m - \alpha^m$ so is inseparable. Note that $\mathrm{char}\, F$ does not divide d if and only if μ_d has order d, so we suppose for the remainder of the proof that this holds.

Any field embedding σ of E in $\overline{F}$ fixing F must send α to $\zeta\alpha$ for some $\zeta \in \mu_d$. If $\mu_d \subset E$, then every such element lies in E, so E/F is Galois. Conversely, if E/F is Galois, then since f does not divide any $x^m - \alpha^m$ with m properly dividing d and μ_d has order d, it has a root of the form $\zeta\alpha$ with $\zeta \in \mu_d$ of order d. Then $\zeta = (\zeta\alpha)\cdot\alpha^{-1} \in E$, so $\mu_d \subset E$.

Finally, if E/F is Galois and $\mu_d \subset F$, then the map χ_a as defined in the statement is bijective by what we have already said, and it satisfies

$$\chi_a(\sigma\tau) = \frac{\sigma\tau(\alpha)}{\alpha} = \frac{\sigma(\alpha)}{\alpha}\cdot\sigma\left(\frac{\tau(\alpha)}{\alpha}\right) = \chi_a(\sigma)\cdot\sigma(\chi_a(\tau)) = \chi_a(\sigma)\chi_a(\tau)$$

for $\sigma, \tau \in \mathrm{Gal}(E/F)$, noting that σ fixes μ_d. □

DEFINITION 10.3.4. Let F be a field containing μ_n for some $n \geq 1$ that is not divisible by char$F$. For any $a \in F^\times$ and extension E/F containing an nth root of a, the *Kummer character* attached to a is the homorphism $\chi_a\colon \mathrm{Gal}(E/F) \to \mu_n$ given by

$$\chi_a(\sigma) = \frac{\sigma(\sqrt[n]{a})}{\sqrt[n]{a}}$$

for $\sigma \in \mathrm{Gal}(E/K)$.

PROPOSITION 10.3.5. *Let F be a field of characteristic not dividing $n \geq 1$, and suppose that F contains the nth roots of unity. Let E be a cyclic extension of F of degree n. Then $E = F(\sqrt[n]{a})$ for some $a \in F^\times$.*

PROOF. Let ζ_n be a primitive nth root of unity in F. Note that $N_{E/F}(\zeta) = \zeta^n = 1$, so Hilbert's Theorem 90 tells us that there exists $\alpha \in E$ and a generator σ of $\mathrm{Gal}(E/F)$ with $\frac{\sigma(\alpha)}{\alpha} = \zeta$. Note that

$$N_{E/F}(\alpha) = \prod_{i=1}^{n}\sigma^i\alpha = \prod_{i=1}^{n}\zeta^i\alpha = \zeta^{\frac{n(n-1)}{2}}\alpha^n = (-1)^{n-1}\alpha^n,$$

so setting $a = -N_{E/F}(-\alpha)$, we have $\alpha^n = a$. Since α has n distinct conjugates in E, we have that $E = F(\alpha)$. □

NOTATION 10.3.6. Let Δ be a subset of a field K, and let $n \geq 1$ be such that K contains the nth roots of unity in $\overline{K}$. Then $K(\sqrt[n]{\Delta})$ is the field given by adjoining an nth root of each element of Δ to K.

THEOREM 10.3.7 (Kummer duality). *Let F be a field of characteristic not dividing $n \geq 1$, and suppose that F contains the nth roots of unity. Let E be a finite abelian extension of F of exponent dividing n, and set $\Delta = E^{\times n} \cap F^\times$. Then $E = F(\sqrt[n]{\Delta})$, and there is a perfect bimultiplicative pairing*

$$\langle\ ,\ \rangle\colon \mathrm{Gal}(E/F) \times \Delta/F^{\times n} \to \mu_n$$

given by $\langle\sigma, a\rangle = \chi_a(\sigma)$ for $\sigma \in \mathrm{Gal}(E/F)$ and $a \in \Delta$.

PROOF. Since $\mu_n \subset F$, Proposition 10.3.3 tells us that the map taking $a \in \Delta$ to its Kummer cocycle χ_a yields an injection

$$\psi\colon \Delta/F^{\times n} \to \mathrm{Hom}(\mathrm{Gal}(E/F), \mu_n).$$

This gives rise to the bimultiplicative Kummer pairing $\langle\ ,\ \rangle$, and it implies that any $a \in \Delta/F^{\times n}$ of order d dividing n pairs with some element of $\mathrm{Gal}(E/F)$ to a dth root of unity.

We claim that ψ is surjective. Let $\chi\colon \mathrm{Gal}(E/F) \to \mu_n$ be a homomorphism, and let $H = \ker\chi$, which by the fundamental theorem of Galois theory corresponds to some cyclic extension K/F of degree dividing n. By Proposition 10.3.5, we have that $K = F(\alpha)$ for some α with $a = \alpha^n \in \Delta$, and then $\chi = \chi_a^k$ for some $k \geq 1$. That is, $\chi = \chi_b$ with $b = a^k \in \Delta$, so $\psi(b) = \chi$. Since $\Delta/F^{\times n}$ is therefore finite of degree $[E:F]$, we have that the map

$$\mathrm{Gal}(E/F) \to \mathrm{Hom}(\Delta/F^{\times n}, \mu_n)$$

induced by the pairing is an isomorphism as well, and thus the Kummer pairing is perfect. □

REMARK 10.3.8. One may replace Δ in Theorem 10.3.7 by any $\Gamma \subseteq \Delta$ with $\Delta = \Gamma K^{\times n}$. Then $\Delta/K^{\times n}$ should be replaced by the isomorphic $\Gamma/(\Gamma \cap K^{\times n})$.

DEFINITION 10.3.9. A finite field extension E/F is *solvable by radicals* if there exists $s \geq 0$ and fields E_i for $0 \leq i \leq s$ with $E_0 = F$, $E \subseteq E_s$, and $E_{i+1} = E_i(\sqrt[n_i]{\alpha_i})$ for some $\alpha_i \in E_i$ and integers $n_i \geq 1$ for $0 \leq i < s$. If we can take $E_s = E$, then we say that E is a *radical extension* of F.

THEOREM 10.3.10. *Let F be a field. Let $f \in F[x]$ be nonconstant, and suppose that its splitting field K has degree over F not divisible by* $\mathrm{char} F$. *Then K is solvable by radicals if and only if* $\mathrm{Gal}(K/F)$ *is a solvable group.*

PROOF. First, take $n = [K:F]$, and consider $L = K(\zeta_n)$ and $E = F(\zeta_n)$ for a primitive nth root of unity ζ_n (which exists by assumption on $\mathrm{char} F$). Then K/F is solvable by radicals if and only if L/E is, since ζ_n is an nth root of 1. Moreover, $L = KE$ is Galois over E, and $\mathrm{Gal}(L/E)$ is solvable if and only if $\mathrm{Gal}(K/F)$ is, since $\mathrm{Gal}(L/E)$ is isomorphic to the subgroup $\mathrm{Gal}(K/K \cap E)$ of $\mathrm{Gal}(K/F)$ by restriction, and $(K \cap E)/F$ is abelian, hence solvable, in that E/F is abelian. Thus, we have reduced to the case that F contains the nth roots of unity.

If K/F is solvable by radicals, then there exists a field L containing F that is a radical extension of F. We claim that we may take L/F to be Galois. Suppose that $K = K_s$ where $K_0 = F$ and $K_{i+1} = K_i(\sqrt[n_i]{\alpha_i})$ with n_i dividing n and $\alpha_i \in K_i$ for $i < s$. Then let $L_0 = F$ and let L_{i+1} be the field given by adjoining to L_i an n_ith root of each conjugate of α_i over F. Then L_{i+1} is the compositum of L_i and the splitting field of the minimal polynomial of $\sqrt[n_i]{\alpha_i}$ over F, so is Galois over F for each i. Moreover, $L = L_s$ is by definition a radical extension of F. Now, $\mathrm{Gal}(L/F)$ is solvable, since $\mathrm{Gal}(L_{i+1}/L_i)$ is abelian in that L_i contains the n_ith root of unity. Since $\mathrm{Gal}(L/F)$ is solvable, so is $\mathrm{Gal}(K/F)$.

Conversely, if $\mathrm{Gal}(K/F)$ is solvable, then we have intermediate fields K_i with $K_0 = F$, $K_s = K$, and $K_i \subset K_{i+1}$ such that K_{i+1}/F is Galois and $\mathrm{Gal}(K_{i+1}/K_i)$ is cyclic of degree dividing n. But then K_{i+1} is a Kummer extension of K_i, given by adjoining the n_ith root of some $\alpha_i \in K_i$, where $n_i = [K_{i+1}:K_i]$. So, K/F is in fact a radical extension. □

COROLLARY 10.3.11. *If F is a field of characteristic not dividing* 6 *and K is the splitting field over F of a polynomial of degree at most* 4, *then K/F is solvable by radicals.*

PROOF. We know that $\mathrm{Gal}(K/F)$ is isomorphic to a subgroup of S_n for n equal to the degree of the polynomial defining K, and S_n is solvable for $n \leq 4$, so $\mathrm{Gal}(K/F)$ is solvable as well. □

EXAMPLE 10.3.12. The splitting field K of the polynomial $f = 2x^5 - 10x + 5 \in \mathbb{Q}[x]$ has Galois group isomorphic to S_5, and S_5 is insolvable, so $K/\mathbb{Q}$ is not solvable by radicals. To see this, note first that the polynomial is irreducible by the Eisenstein criterion for the prime 5. So, 5 divides $[K : \mathbb{Q}]$, and hence the image G in S_5 of $\mathrm{Gal}(K/\mathbb{Q})$ under a permutation representation of the roots contains a 5 cycle. Moreover, $f' = 10(x^4 - 1)$ has real roots at ± 1 and $f(-1) > 0$ while $f(1) < 0$, so f has exactly three real roots. In particular, if $\tau \in \mathrm{Gal}(K/\mathbb{Q})$ is the restriction of complex conjugation, then τ fixes the three real roots and transposes the two imaginary roots, so G contains a transposition. But S_5 is generated by any five cycle and any transposition, so $\mathrm{Gal}(K/\mathbb{Q}) \cong G = S_5$.

10.4. Linearly disjoint extensions

PROPOSITION 10.4.1. *Let K be a field, and let $f \in K[x]$ be monic and irreducible. Let M be a field extension of K, and suppose that f factors as $\prod_{i=1}^m f_i^{e_i}$ in $M[x]$, where the f_i are irreducible and distinct and each e_i is positive. Then we have an isomorphism*

$$\kappa \colon K[x]/(f) \otimes_K M \xrightarrow{\sim} \prod_{i=1}^m M[x]/(f_i^{e_i})$$

of M-algebras such that if $g \in K[x]$, then $\kappa((g + (f)) \otimes 1) = (g + (f_i^{e_i}))_i$.

PROOF. Note that we have a canonical isomorphism $K[x] \otimes_K M \xrightarrow{\sim} M[x]$ that gives rise to the first map in the composition

$$K[x]/(f) \otimes_K M \xrightarrow{\sim} M[x]/(f) \xrightarrow{\sim} \prod_{i=1}^m M[x]/(f_i^{e_i}),$$

the second isomorphism being the Chinese remainder theorem. The composition is κ. □

We have the following consequence.

LEMMA 10.4.2. *Let L/K be a finite separable extension of fields, and let M be an algebraically closed field containing K. Then we have an isomorphism of M-algebras*

$$\kappa \colon L \otimes_K M \xrightarrow{\sim} \prod_{\sigma \colon L \hookrightarrow M} M,$$

where the product is taken over field embeddings of L in M fixing K, such that

$$\kappa(\beta \otimes 1) = (\sigma\beta)_\sigma$$

for all $\beta \in L$.

PROOF. Write $L = K(\theta)$, and let $f \in K[x]$ be the minimal polynomial of θ. Then we define κ as the composition

$$L \otimes_K M \xrightarrow{\sim} \frac{M[x]}{(x - \sigma(\theta))} \xrightarrow{\sim} \prod_{\sigma \colon L \hookrightarrow M} M,$$

where the first isomorphism is that of Proposition 10.4.1 and the second takes x to $\sigma(\theta)$ in the coordinate corresponding to σ. Any $\beta \in L$ has the form $g(\theta)$ for some $g \in K[x]$, and since any $\sigma \colon L \hookrightarrow M$ fixing K fixes the coefficients of g, we have $\kappa(\beta \otimes 1)$ is as stated. □

REMARK 10.4.3. If we compose κ of Lemma 10.4.2 with the natural embedding $L \hookrightarrow L \otimes_K M$ that takes $\alpha \in L$ to $\alpha \otimes 1$, then the composition

$$\iota_M \colon L \to \prod_{\sigma \colon L \hookrightarrow M} M$$

is the product of the field embeddings σ of L in M fixing K.

DEFINITION 10.4.4. Let K be a field and L and M be extensions of K both contained in some field Ω. We say that L and M are *linearly disjoint* over K if every K-linearly independent subset of L is M-linearly independent.

LEMMA 10.4.5. *Let K be a field and L and M be extensions of K both contained in some field Ω. If L and M are linearly disjoint over K, then $L \cap M = K$.*

PROOF. If $x \in L \cap M$ with $x \notin K$, then x and 1 are elements of L that are K-linearly independent but not M-linearly independent, so L and M are not linearly disjoint over K. □

From the definition, it may not be clear that the notion of linear disjointness is a symmetric one. However, this follows from the following.

PROPOSITION 10.4.6. *Let K be a field and L and M be extensions of K both contained in some field Ω. Then L and M are linearly disjoint over K if and only if the map $\varphi \colon L \otimes_K M \to LM$ induced by multiplication is an injection.*

PROOF. Suppose that $\gamma_1, \ldots, \gamma_s \in M$ are L-linearly dependent, and write $\sum_{i=1}^s \beta_i \gamma_i = 0$ for some $\beta_i \in L$. If φ is injective, then we must have $\sum_{i=1}^s \beta_i \otimes \gamma_i = 0$, which means that the γ_i are K-linearly dependent.

Conversely, let L and M be linearly disjoint over K. Suppose that we have a nonzero

$$x = \sum_{i=1}^{s} \beta_i \otimes \gamma_i \in \ker \varphi$$

for some $\beta_i \in L$ and $\gamma_i \in M$, with s taken to be minimal. If $x \neq 0$, then the γ_i are L-linearly dependent, so they are K-linearly dependent. In this case, without loss of generality, we may suppose that

$$\gamma_s + \sum_{i=1}^{s-1} \alpha_i \gamma_i = 0$$

for some α_i in K. Then

$$x = \sum_{i=0}^{s-1} (\beta_i - \alpha_i \beta_s) \otimes \gamma_i,$$

contradicting minimality. Thus $\ker \varphi = 0$. □

COROLLARY 10.4.7. *Let K be a field and L and M be extensions of K both contained in a given algebraic closure of K. Then L and M are linearly disjoint over K if and only if $L \otimes_K M$ is a field.*

PROOF. Note that LM is a union of subfields of the form $K(\alpha,\beta)$ with $\alpha \in L$ and $\beta \in M$. Since α and β are algebraic over K, we have $K(\alpha,\beta) = K[\alpha,\beta]$, and every element of the latter ring is a K-linear combination of monomials in α and β. Thus φ of Proposition 10.4.6 is surjective, and the result follows from the latter proposition. □

COROLLARY 10.4.8. *Let K be a field and L and M be finite extensions of K inside a given algebraic closure of K. Then $[LM : K] = [L : K][M : K]$ if and only if L and M are linearly disjoint over K.*

PROOF. Again, we have the surjection $\varphi\colon L \otimes_K M \to LM$ given by multiplication which is an injection if and only if L and M are linearly disjoint by Proposition 10.4.6. As $L \otimes_K M$ has dimension $[L : K][M : K]$ over K, the result follows. □

REMARK 10.4.9. Suppose that $L = K(\theta)$ is a finite extension of K. To say that L is linearly disjoint from a field extension M of K is by Propostion 10.4.1 exactly to say that the minimal polynomial of θ in $K[x]$ remains irreducible in $M[x]$.

We prove the following in somewhat less generality than possible.

LEMMA 10.4.10. *Let L be a finite Galois extension of a field K inside an algebraic closure Ω of K, and let M be an extension of K in Ω. Then L and M are linearly disjoint if and only if $L \cap M = K$.*

PROOF. We write $L = K(\theta)$ for some $\theta \in L$, and let $f \in K[x]$ be the minimal polynomial of θ. As $\mathrm{Gal}(LM/M) \cong \mathrm{Gal}(L/(L\cap M))$ by restriction, we have $L \cap M = K$ if and only if $[LM : M] = [L : K]$. Since $LM = M(\theta)$, this occurs if and only if f is irreducible in $M[x]$. The result then follows from Remark 10.4.9. □

10.5. Normal bases

DEFINITION 10.5.1. A *normal basis* of a finite Galois extension L/K is a basis of L as a K-vector space of the form $\{\sigma(\alpha) \mid \sigma \in \mathrm{Gal}(L/K)\}$ for some $\alpha \in L$.

The goal of this section is to prove E. Noether's theorem that every finite Galois extension has a normal basis. We start with the following lemma.

LEMMA 10.5.2. *Let L/K be a finite Galois extension with Galois group $\{\sigma_1,\dots,\sigma_n\}$, where $n = [L : K]$. Let $\{\alpha_1,\dots,\alpha_n\}$ be a basis of L as a K-vector space. Then the set*

$$\{(\sigma_1(\alpha_j),\dots,\sigma_n(\alpha_j)) \mid 1 \le j \le n\}$$

is an L-basis of L^n.

PROOF. Let W be the L-span of the subset of L^n in question. Set $W^\vee = \mathrm{Hom}_L(W,L)$, and let $\varphi \in (L^n)^\vee$ be such that $\varphi(W) = 0$. It suffices to show that $\varphi = 0$. Note that there exists $u = (a_1,\dots,a_n) \in L^n$ such that $\varphi(v) = u^T v$ for all $v \in L^n$, so $\sum_{i=1}^n a_i\sigma_i(\alpha_j) = 0$ for all $1 \le j \le n$. As $\{\alpha_1,\dots,\alpha_n\}$ is a K-basis of L, we therefore have that $\sum_{i=1}^n a_i\sigma_i$ vanishes on L. Since the σ_i are L-linearly independent, we have $a_i = 0$ for all i, and therefore $\varphi = 0$. □

LEMMA 10.5.3. *Every finite cyclic extension of fields has a normal basis.*

PROOF. Let L/K be finite cyclic of degree n, generated by an element σ. Then $K[\mathrm{Gal}(L/K)]$ is isomorphic to $K[x]/(x^n-1)$ via the unique K-algebra homomorphism that takes σ to x. As L is a $K[\mathrm{Gal}(L/K)]$-module, it becomes a $K[x]$-module annihilated by x^n-1. If $f=\sum_{i=0}^{n-1}c_ix^i\in K[x]$ annihilates L, then $\sum_{i=0}^{n-1}c_i\sigma^i(\alpha)=0$ for all $\alpha\in L$, which by the linear independence of the σ^i forces f to be zero. Thus, the annihilator of L is (x^n-1), and by the structure theorem for finitely generated modules over the PID $K[x]$, this means that L has a $K[x]$-summand isomorphic to $K[x]/(x^n-1)$, generated by some $\alpha\in L$. Since the latter module has K-dimension n, as does L, the elements $\{\alpha,\sigma(\alpha),\dots,\sigma^{n-1}(\alpha)\}$ form a K-basis of L.

□

THEOREM 10.5.4 (Normal basis theorem). *Every finite Galois extension of fields has a normal basis.*

PROOF. Let L/K be a finite Galois extension of degree n. Since any finite extension of finite fields is cyclic, we may by Lemma 10.5.3 suppose that K is infinite. Write $\mathrm{Gal}(L/K)=\{\sigma_1,\dots,\sigma_n\}$ and $\sigma_1=1$. Let $\{\alpha_1,\dots,\alpha_n\}$ be a basis of L as a K-vector space. It suffices to find $\beta\in L$ with $\mathrm{D}(\sigma_1(\beta),\dots,\sigma_n(\beta))\neq 0$ by Corollary 10.2.16.

Define an element $p\in L[x_1,\dots,x_n]$ by

$$p(x_1,\dots,x_n)=\det\left(\sum_{k=1}^{n}\sigma_j^{-1}\sigma_i(\alpha_k)x_k\right)^2.$$

Note that the coefficients of p are fixed by the elements of $\mathrm{Gal}(L/K)$, since they permute the columns of the matrix. By Lemma 10.5.2, we can find $\beta_j\in L$ for $1\le j\le n$ be such that

$$\sum_{j=1}^{n}\beta_j(\sigma_1(\alpha_j),\sigma_2(\alpha_j),\dots,\sigma_n(\alpha_j))=(1,0,\dots,0).$$

Then for all $1\le i,j\le n$, we have

$$\sum_{k=1}^{n}\sigma_j^{-1}\sigma_i(\alpha_k)\beta_k=\delta_{i,j},$$

so $p(\beta_1,\dots,\beta_n)=\det(I_n)^2=1$, so $p\neq 0$. Since K is infinite, there exist $a_1,\dots,a_n\in K$ with $p(a_1,\dots,a_n)\neq 0$. For $\gamma=\sum_{j=1}^{n}a_i\alpha_i$, we have by Proposition 10.2.9 the first equality in

$$\mathrm{D}(\sigma_1(\gamma),\dots,\sigma_n(\gamma))=\det(\sigma_j^{-1}\sigma_i(\gamma))^2=p(a_1,\dots,a_n)\neq 0.$$

□

10.6. Profinite groups

DEFINITION 10.6.1. A *topological group* G is a group endowed with a topology with respect to which both the multiplication map $G\times G\to G$ and the inversion map $G\to G$ that takes an element to its inverse are continuous.

EXAMPLES 10.6.2.

a. The groups $\mathbb{R}$, $\mathbb{C}$, $\mathbb{R}^\times$, and $\mathbb{C}^\times$ are continuous with respect to the topologies defined by their absolute values.

b. Any group can be made a topological group by endowing it with the discrete topology.

REMARK 10.6.3. We may consider the category of topological groups, in which the maps are continuous homomorphisms between topological groups.

DEFINITION 10.6.4. A homomorphism $\phi \colon G \to G'$ between topological groups G and G' is a *topological isomorphism* if it is both an isomorphism and a homeomorphism.

The following lemma is almost immediate, since elements of a group are invertible.

LEMMA 10.6.5. *Let G be a topological group and $g \in G$. Then the map $m_g \colon G \to G$ with $m_g(a) = ga$ for all $a \in G$ is a topological isomorphism.*

We also have the following.

LEMMA 10.6.6. *A group homomorphism $\phi \colon G \to G'$ between topological groups is continuous if and of only, for each open neighborhood U of* 1 *in G' with $1 \in U$, the set $\phi^{-1}(U)$ contains an open neighborhood of* 1.

PROOF. We consider the non-obvious direction. Let V be an open set in G', and suppose that $g \in G$ is such that $h = \phi(g) \in V$. Then $h^{-1}V$ is open in G' as well, by Lemma 10.6.5. By assumption, there exists an open neighborhood W of 1 in G contained in $\phi^{-1}(h^{-1}V)$, and so gW is an open neighborhood of g in G such that $\phi(gW) \subseteq V$. Hence, ϕ is continuous. □

LEMMA 10.6.7. *Let G be a topological group.*

a. Any open subgroup of G is closed.

b. Any closed subgroup of finite index in G is open.

PROOF. If H is an open (resp., closed) subgroup of G, then its cosets are open (resp., closed) as well. Moreover, $G - H$ is the union of the nontrivial cosets of H. Therefore, $G - H$ is open if G is open and closed if G is closed of finite index, so that there are only finitely many cosets of H. □

LEMMA 10.6.8. *Every open subgroup of a compact group G is of finite index in G.*

PROOF. Let H be a open subgroup of G. Note that G is the union of its distinct H-cosets, which are open and disjoint. Since G is compact, there can therefore only be finitely many cosets, which is to say that H is of finite index in G. □

We leave it to the reader to verify the following.

LEMMA 10.6.9.

a. A subgroup of a topological group is a topological group with respect to the subspace topology.

b. The quotient of a topological group G by a normal subgroup N is a topological group with respect to the quotient topology, and it is Hausdorff if G is Hausdorff and N is closed.

c. A direct product of topological groups is a topological group with respect to the product topology.

REMARK 10.6.10. The category of topological Hausdorff abelian groups is not abelian, though it is additive and admits kernels and cokernels. For instance, consider the inclusion map $\iota\colon \mathbb{Q} \to \mathbb{R}$ with $\mathbb{R}$ having its usual topology and $\mathbb{Q}$ having the subspace topology. Then $\ker \iota = 0$ and $\operatorname{coker} \iota = 0$ (since $\mathbb{Q}$ is dense in $\mathbb{R}$, and thus every continuous map from $\mathbb{R}$ is determined by its values on $\mathbb{Q}$). By Proposition 8.8.15, we have $\operatorname{im} \iota \cong \mathbb{R}$ but $\operatorname{coim} \iota \cong \mathbb{Q}$.

Recall the definitions of a directed set, inverse system, and the inverse limit.

DEFINITION 10.6.11. A *directed set* $I = (I, \geq)$ is a partially ordered set such that for every $i, j \in I$, there exists $k \in I$ with $k \geq i$ and $k \geq j$.

DEFINITION 10.6.12. Let I be a directed set. An *inverse system* $(G_i, \phi_{i,j})$ of groups over the indexing set I is a set

$$\{G_i \mid i \in I\}$$

of groups and a set

$$\{\phi_{i,j}\colon G_i \to G_j \mid i, j \in I, i \geq j\}$$

of group homomrphisms.

DEFINITION 10.6.13. An *inverse limit*

$$G = \varprojlim_i G_i$$

of an inverse system of groups $(G_i, \phi_{i,j})$ over a directed indexing set I is a pair $G = (G, \{\pi_i \mid i \in I\})$ consisting of a group G and homomorphisms $\pi_i\colon G \to G_i$ such that $\phi_{i,j} \circ \pi_i = \pi_j$ for all $i, j \in I$ with $i \geq j$ that satisfy the following universal property: Given a group G' and maps $\pi_i'\colon G' \to G_i$ for $i \in I$ such that $\phi_{i,j} \circ \pi_i' = \pi_j'$ for all $i \geq j$, there exists a unique map $\psi\colon G' \to G$ such that $\pi_i' = \pi_i \circ \psi$ for all $i \in I$.

By the universal property, any two inverse limits of an inverse system of groups are canonically isomorphic (via compatible maps).

REMARK 10.6.14. We may make the latter definition more generally with any category $\mathscr{C}$ replacing the category of groups. The groups are replaced with objects in $\mathscr{C}$ and the group homomorphisms with morphisms in $\mathscr{C}$. Moreover, we may view the system of groups as a covariant functor to the category $\mathscr{C}$ from the category that has the elements of I as its objects and morphisms $i \to j$ for each $i, j \in I$ with $i \leq j$.

We may give a direct construction of an inverse llimit of an inverse system of groups as follows. The proof is left to the reader.

PROPOSITION 10.6.15. *Let $(G_i, \phi_{i,j})$ be an inverse system of groups over an indexing set I. Then the an inverse limit of the system is given explicitly by the group*

$$G = \left\{ (g_i)_i \in \prod_{i \in I} G_i \mid \phi_{i,j}(g_i) = g_j \right\}$$

and the maps $\pi_i\colon G \to G_i$ for $i \in I$ that are the compositions of the $G \to \prod_{i\in I} G_i \to G_i$ of inclusion followed by projection.

We may endow an inverse limit of groups with a topology as follows.

DEFINITION 10.6.16. Let $(G_i, \phi_{i,j})$ be an inverse system of topological groups over an indexing set I, with continuous maps. Then the *inverse limit topology* on the inverse limit G of Proposition 10.6.15 is the subspace topology for the product topology on $\prod_{i\in I} G_i$.

LEMMA 10.6.17. *The inverse limit of an inverse system $(G_i, \phi_{i,j})$ of topological groups (over a directed indexing set I) is a topological group under the inverse limit topology.*

PROOF. The maps

$$\prod_{i\in I} G_i \times \prod_{i\in I} G_i \to \prod_{i\in I} G_i \quad \text{and} \quad \prod_{i\in I} G_i \to \prod_{i\in I} G_i$$

given by componentwise multiplication and inversion are clearly continuous, and this continuity is preserved under the subspace topology on the inverse limit. □

REMARK 10.6.18. In fact, the inverse limit of an inverse system of topological groups and continuous maps, when endowed with the product topology, is an inverse limit in the category of topological groups.

When we wish to view it as a topological group, we typically endow a finite group with the discrete topology.

DEFINITION 10.6.19. A *profinite group* is an inverse limit of a system of finite groups, endowed with the inverse limit topology for the discrete topology on the finite groups.

Recall the following definition.

DEFINITION 10.6.20. A topological space is *totally disconnected* if and only if every point is a connected component.

We leave the following as difficult exercises.

PROPOSITION 10.6.21. *A compact Hausdorff space is totally disconnected if and only if it has a basis of open neighborhoods that are also closed.*

PROPOSITION 10.6.22. *A compact Hausdorff group that is totally disconnected has a basis of neighborhoods of* 1 *consisting of open normal subgroups (of finite index).*

We may now give a topological characterization of profinite groups.

THEOREM 10.6.23. *A profinite topological group G is compact, Hausdorff, and totally disconnected.*

PROOF. First, suppose that G is profinite, equal to an inverse limit of a system $(G_i, \phi_{i,j})$ of finite groups over an indexing set I. The direct product $\prod_{i\in I} G_i$ of finite (discrete) groups G_i is compact Hausdorff (compactness being Tychonoff's theorem). As a subset of the direct product, G is Hausdorff, and to see it is compact, we show that G is closed. Suppose that

$$(g_i)_i \in \prod_{i \in I} G_i$$

with $(g_i)_i \notin G$, and choose $i, j \in I$ with $i > j$ and $\phi_{i,j}(g_i) \neq g_j$. The open subset

$$\left\{ (h_k)_k \in \prod_{k \in I} G_k \mid h_i = g_i, h_j = g_j \right\}$$

of the direct product contains $(g_i)_i$ and has trivial intersection with G. In that the complement of G is open, G itself is closed. Finally, note that any open set $\prod_{i \in I} U_i$ with each U_i open in G_i (i.e., an arbitrary subset) and $U_i = G_i$ for all but finitely many i is also closed. That is, its complement is the intersection

$$\bigcap_{j \in I} \left((G_j - U_j) \times \prod_{i \in I - \{j\}} U_i \right)$$

of open sets, which is actually equal to the finite intersection over $j \in I$ with $U_i \neq G_i$. It is therefore open, and by Proposition 10.6.21, the group G is totally disconnected. □

REMARK 10.6.24. We leave it to the reader to check that the converse to Theorem 10.6.23 also holds. They key is found in the proof of part a of the following proposition.

PROPOSITION 10.6.25. *Let G be a profinite group, and let $\mathscr{U}$ be the set of all open normal subgroups of G. Then the following canonical homomorphisms are homeomorphisms:*

a. $G \to \varprojlim_{N \in \mathscr{U}} G/N$,

b. $H \to \varprojlim_{N \in \mathscr{U}} H/(H \cap N)$, for H a closed subgroup of G, and

c. $G/K \to \varprojlim_{N \in \mathscr{U}} G/NK$, for K a closed normal subgroup of G.

PROOF. We prove part a. The continuous map ϕ from G to the inverse limit Q of its quotients has closed image, and ϕ is injective since $\mathscr{U}$ is a basis of 1 in G as in Proposition 10.6.22. Suppose that $(g_N N)_{N \in \mathscr{U}}$ is not in the image of ϕ, which is exactly to say that the intersection of the closed sets $g_N N$ is empty. Since G is compact this implies that some finite subset of the $\{g_N N \mid N \in \mathscr{U}\}$ is empty, and letting M be the intersection of the N in this subset, we see that $g_M M = \varnothing$, which is a contradiction. In other words, ϕ is surjective. □

The following is a consequence of Proposition 10.6.25a. We leave the proof to the reader.

COROLLARY 10.6.26. *Let G be a profinite group and $\mathscr{V}$ a set of open normal subgroups of G that forms a basis of open neighborhoods of* 1. *Then the homomorphism*

$$G \to \varprojlim_{N \in \mathscr{V}} G/N$$

is a homeomorphism.

The following lemma will be useful later.

LEMMA 10.6.27. *The closed subgroups of a profinite group are exactly those that may be written as intersections of open subgroups.*

PROOF. In a topological group, an open subgroup is also closed, an arbitrary intersection of closed sets is closed, and an arbitrary intersection of subgroups is a subgroup, so an intersection of open subgroups is a closed subgroup. Let $\mathscr{U}$ denote the set of open subgroups of a profinite group G. Let H be a closed subgroup of G. It follows from Proposition 10.6.25b and the second isomorphism theorem that the set of subgroups of the norm NH with N open normal in G has intersection H. Note that each NH is open as a union of open subgroups, so it is open. □

We may also speak of pro-p groups.

DEFINITION 10.6.28. A *pro-p group*, for a prime p, is an inverse limit of a system of finite p-groups.

We may also speak of profinite and pro-p completions of groups.

DEFINITION 10.6.29. Let G be a group.

a. The *profinite completion* $\hat{G}$ of G is the inverse limit of its finite quotients G/N, for N a normal subgroup of finite index in G, together with the natural quotient maps $G/N \to G/N'$ for $N \le N'$.

b. The *pro-p completion* $G^{(p)}$ of G, for a prime p, is the inverse limit of the finite quotients of G of p-power order, i.e., of the G/N for $N \trianglelefteq G$ with $[G:N]$ a power of p, together with the natural quotient maps.

REMARK 10.6.30. A group G is endowed with a canonical homomorphism to its profinite completion $\hat{G}$ by the universal property of the inverse limit.

REMARK 10.6.31. We may also speak of topological rings and fields, where multiplication, addition, and the additive inverse map are continuous, and in the case of a topological field, the multiplicative inverse map on the multiplicative group is continuous as well. We may speak of profinite rings as inverse limits by quotients by two-sided ideals of finite index (or for pro-p rings, of p-power index).

The next proposition shows that $\mathbb{Z}_p$ is the pro-p completion of $\mathbb{Z}$.

PROPOSITION 10.6.32. *Let p be a prime. We have an isomorphism of rings*

$$\psi\colon \mathbb{Z}_p \xrightarrow{\sim} \varprojlim_{k\ge 1} \mathbb{Z}/p^k\mathbb{Z}, \qquad \sum_{i=0}^{\infty} a_i p^i \mapsto \left(\sum_{i=0}^{k-1} a_i p^i\right)_k,$$

where the maps $\mathbb{Z}/p^{k+1}\mathbb{Z} \to \mathbb{Z}/p^k\mathbb{Z}$ in the system are the natural quotient maps. Moreover, ψ is a homeomorphism.

PROOF. The canonical quotient map $\psi_k\colon \mathbb{Z}_p \to \mathbb{Z}/p^k\mathbb{Z}$ is the kth coordinate of ψ, which is then a ring homomorphism by the universal property of the inverse limit. The kernel ψ is the intersection of the kernels of the maps ψ_k, which is exactly

$$\bigcap_k p^k\mathbb{Z}_p = 0.$$

Moreover, any sequence of partial sums modulo increasing powers of p has a limit in $\mathbb{Z}_p$, which maps to the sequence under ψ. The open neighborhood $p^n\mathbb{Z}_p$ of 0 in the p-adic topology is sent to the intersection

$$\left(\prod_{k=1}^{n}\{0\} \times \prod_{k=n+1}^{\infty} \mathbb{Z}_p/p^k\mathbb{Z}_p\right) \cap \left(\varprojlim_{k\geq 1} \mathbb{Z}/p^k\mathbb{Z}\right),$$

which is open in the product topology. On the other hand, the inverse image of a basis open neighborhood

$$\left(\prod_{k=1}^{n} U_k \times \prod_{k=n+1}^{\infty} \mathbb{Z}_p/p^k\mathbb{Z}_p\right) \cap \left(\varprojlim_{k\geq 1} \mathbb{Z}/p^k\mathbb{Z}\right)$$

with $0 \in U_k$ for all $1 \leq k \leq n$ under ψ clearly contains $p^n\mathbb{Z}_p$. It then follows from Lemma 10.6.6 that ψ is a homeomorphism. □

DEFINITION 10.6.33. The *Prüfer ring* $\hat{\mathbb{Z}}$ is the profinite completion of $\mathbb{Z}$. That is, we have

$$\mathbb{Z} \cong \varprojlim_{n\geq 1} \mathbb{Z}/n\mathbb{Z}$$

with respect to the quotient maps $\mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ for $m \mid n$.

Since $\mathbb{Z}/n\mathbb{Z}$ may be written as a direct product of the $\mathbb{Z}/p^k\mathbb{Z}$ for primes p with p^k exactly dividing n, we have the following.

LEMMA 10.6.34. *We have an isomorphism of topological rings*

$$\hat{\mathbb{Z}} \cong \prod_{p \text{ prime}} \mathbb{Z}_p.$$

EXAMPLE 10.6.35. The free profinite (or pro-p) group on a generating set S is the profinite (resp., pro-p) completion of the free group on S.

REMARK 10.6.36. As with free groups, closed subgroups of free profinite (or pro-p) groups are free profinite (or pro-p) groups. Moreover, every profinite (resp., pro-p) group is a topological quotient of the free group on a set of its generators, so we may present such groups via generators and relations much as before.

DEFINITION 10.6.37. A subset S of a topological group G is said to be a *topological generating set* of G if G is the closure of the subgroup generated by S.

DEFINITION 10.6.38. We say that a topological group is *(topologically) finitely generated* if it has a finite set of topological generators.

REMARK 10.6.39. If G is a free profinite (or pro-p) group on a set S, then it is topologically generated by S.

We leave a proof of the following to the reader.

LEMMA 10.6.40. *Let G be a topological group, and let H be a (normal) subgroup. Then the closure $\overline{H}$ of H is also a (normal) subgroup of G.*

10.7. Infinite Galois theory

Recall that an algebraic extension of fields L/K is Galois if it is normal, so that every polynomial in $K[x]$ that has a root in L splits completely, and separable, so that no irreducible polynomial in $K[x]$ has a double root in L. The Galois group $\mathrm{Gal}(L/K)$ of such an extension is the group of automorphisms of L that fix K.

In the setting of finite Galois extensions L/K, the subfields E of L containing F are in one-to-one correspondence with the subgroups H of $\mathrm{Gal}(L/K)$. In fact, the maps $E \mapsto \mathrm{Gal}(L/E)$ and $H \mapsto L^H$ give inverse bijections between these sets. This is not so in the setting of infinite Galois extensions, where there are rather more subgroups than there are subfields. To fix this issue, we place a topology on $\mathrm{Gal}(L/K)$ and consider only the closed subgroups under this topology. The above-described correspondences then work exactly as before.

PROPOSITION 10.7.1. *Let L/K be a Galois extension of fields. Let $\mathscr{E}$ denote the set of finite Galois extensions of K contained in L, ordered by inclusion. This is a directed set. Let ρ be the map*

$$\rho \colon \mathrm{Gal}(L/K) \to \varprojlim_{E\in\mathscr{E}} \mathrm{Gal}(E/K)$$

defined by the universal property of the inverse limit, with the maps $\mathrm{Gal}(E'/K) \to \mathrm{Gal}(E/K)$ *for $E, E' \in \mathscr{E}$ with $E \subseteq E'$ and the maps* $\mathrm{Gal}(L/K) \to \mathrm{Gal}(E/K)$ *for $E \in \mathscr{E}$ being restriction maps. Then ρ is an isomorphism.*

PROOF. Let $\sigma \in \mathrm{Gal}(L/K)$. If $\sigma|_E = 1$ for all $E \in \mathscr{E}$, then since

$$L = \bigcup_{E\in\mathscr{E}} E,$$

we have that $\sigma = 1$. On the other hand, if elements $\sigma_E \in \mathrm{Gal}(E/K)$ for each $E \in \mathscr{E}$ are compatible under restriction, then define $\sigma \in \mathrm{Gal}(L/K)$ by $\sigma(\alpha) = \sigma_E(\alpha)$ if $\alpha \in E$. Then, if $\alpha \in E'$ for some $E' \in \mathscr{E}$ as well, then

$$\sigma_{E'}(\alpha) = \sigma_{E\cap E'}(\alpha) = \sigma_E(\alpha),$$

noting that $E \cap E' \in \mathscr{E}$. Therefore, σ is well-defined, and so ρ is bijective. □

Proposition 10.7.1 gives us an obvious topology to place on the Galois group of a Galois extension.

DEFINITION 10.7.2. Let L/K be a Galois extension of fields. The *Krull topology* on $\mathrm{Gal}(L/K)$ is the unique topology under which the set of $\mathrm{Gal}(L/E)$ for E/K finite Galois with $E \subseteq L$ forms a basis of open neighborhoods of 1.

REMARK 10.7.3. The Krull topology agrees with the inverse limit topology induced by the isomorphism of Proposition 10.7.1, since

$$1 \to \mathrm{Gal}(L/E) \to \mathrm{Gal}(L/K) \to \mathrm{Gal}(E/K) \to 1$$

is exact. Therefore, if L/K is Galois, then $\mathrm{Gal}(L/K)$ is a topological group under the Krull topology.

LEMMA 10.7.4. *Let L/K be a Galois extension of fields. The open subgroups in* $\mathrm{Gal}(L/K)$ *are exactly those subgroups of the form* $\mathrm{Gal}(L/E)$ *with E an intermediate field in L/K of finite degree over K.*

PROOF. First, let E be an intermediate field in L/K of finite degree. Let E' be the Galois closure of E in L, which is of finite degree over K. Then $\mathrm{Gal}(L/E')$ is an open normal subgroup under the Krull topology, contained in $\mathrm{Gal}(L/E)$. Since $\mathrm{Gal}(L/E)$ is then a union of left $\mathrm{Gal}(L/E')$-cosets, which are open, we have that $\mathrm{Gal}(L/E)$ is open.

Conversely, let H be an open subgroup in $\mathrm{Gal}(L/K)$. Then H contains $\mathrm{Gal}(L/E)$ for some finite Galois extension E/K in L. Any $\alpha \in L^H$, where L^H is the fixed field of H in L, is contained in $M^{\mathrm{Gal}(L/E)}$, where M is the Galois closure of $K(\alpha)$. Since the restriction map $\mathrm{Gal}(L/E) \to \mathrm{Gal}(M/E)$ is surjective, we then have $\alpha \in M^{\mathrm{Gal}(M/E)}$. But M/K is finite, so $M^{\mathrm{Gal}(M/E)} = E$ by the fundamental theorem of Galois theory. Thus $L^H \subseteq E$.

Let $\bar{H}$ be the image of H under the restriction map $\pi\colon \mathrm{Gal}(L/K) \to \mathrm{Gal}(E/K)$. As $\mathrm{Gal}(L/E) \leq H$, we have that $\pi^{-1}(\bar{H}) = H$. We remark that $\bar{H} = \mathrm{Gal}(E/L^H)$, since $\bar{H} = \mathrm{Gal}(E/E^{\bar{H}})$ by the fundamental theorem of Galois theory for finite extensions and $L^H = E^H = E^{\bar{H}}$. But $\pi^{-1}(\bar{H})$ is then $\mathrm{Gal}(L/L^H)$ as well. □

From this, we may derive the following.

LEMMA 10.7.5. *Let L/K be a Galois extension of fields. The closed subgroups of* $\mathrm{Gal}(L/K)$ *are exactly those of the form* $\mathrm{Gal}(L/E)$ *for some intermediate field E in the extension L/K.*

PROOF. Under the Krull topology on $\mathrm{Gal}(L/K)$, the open subgroups are those of the form $\mathrm{Gal}(L/E)$ with E/K finite. By Lemma 10.6.27, we have therefore that the closed subgroups are those that are intersections of $\mathrm{Gal}(L/E)$ over a set S of finite degree over K intermediate fields E. Any such intersection necessarily fixes the compositum $E' = \prod_{E \in S} E$, while if an element of $\mathrm{Gal}(L/K)$ fixes E', then it fixes every $E \in S$, so lies in the intersection. That is, any closed subgroup has the form

$$\mathrm{Gal}(L/E') = \bigcap_{E \in S} \mathrm{Gal}(L/E).$$

□

THEOREM 10.7.6 (Fundamental theorem of Galois theory). *Let L/K be a Galois extension. Then there are inverse one-to-one, inclusion reversing correspondences*

$$\{\text{intermediate extensions in } L/K\} \underset{\theta}{\overset{\psi}{\rightleftarrows}} \{\text{closed subgroups of } \mathrm{Gal}(L/K)\}$$

given by $\psi(E) = \mathrm{Gal}(L/E)$ for any intermediate extension E in L/K and $\theta(H) = L^H$ for any closed subgroup H of $\mathrm{Gal}(L/K)$*. These correspondences restrict to bijections between the normal extensions of K in L and the closed normal subgroups of* $\mathrm{Gal}(L/K)$*, as well as to bijections between the finite degree (normal) extensions of K in L and the open (normal) subgroups of* $\mathrm{Gal}(L/K)$*. For any E of finite degree and the corresponding closed of finite index H, we have*

$$[L:E] = \mathrm{Gal}(L/E) \quad \text{and} \quad |H| = [L:L^H].$$

Moreover, if E is normal over K (resp., $H \trianglelefteq \mathrm{Gal}(L/K)$ is closed), then restriction induces a topological isomorphism

$$\mathrm{Gal}(L/K)/\mathrm{Gal}(L/E) \xrightarrow{\sim} \mathrm{Gal}(E/K)$$

(resp., $\mathrm{Gal}(L/K)/H \xrightarrow{\sim} \mathrm{Gal}(L^H/K)$*).*

PROOF. We will derive this from the fundamental theorem of Galois theory for finite Galois extensions. Let E be an intermediate extension in L/K. Then $E \subseteq L^{\mathrm{Gal}(L/E)}$ by definition. Let $x \in L^{\mathrm{Gal}(L/E)}$. The Galois closure M of $E(x)$ in L is of finite degree over E. But every element of $\mathrm{Gal}(M/E)$ extends to an element of $\mathrm{Gal}(L/E)$, which fixes x. So $x \in M^{\mathrm{Gal}(M/E)}$, which equals E by fundamental theorem of Galois theory for finite Galois extensions. Since x was arbitrary, we have $E = L^{\mathrm{Gal}(L/E)}$. In other words, $\theta(\psi(E)) = E$.

Let H be a closed subgroup of $\mathrm{Gal}(L/K)$. In Lemma 10.7.5, we saw that $H = \mathrm{Gal}(L/E)$ for some intermediate E in L/K. Since $E = L^{\mathrm{Gal}(L/E)} = L^H$ from what we have shown, we have that $H = \mathrm{Gal}(L/L^H)$. Therefore, $\psi(\theta(H)) = H$. It follows that we have the desired inclusion-reserving one-to-one correspondences. The other claims are then easily checked, or follow from the case of finite degree, and are left to the reader. □

DEFINITION 10.7.7. A *separable closure* of a field L is any field that contains all roots of all separable polynomials in L.

NOTATION 10.7.8. We typically denote a separable closure of L by L^{sep}.

REMARK 10.7.9. If one fixes an algebraically closed field Ω containing L, then there is a unique separable closure of L in Ω, being the subfield generated by the roots of all separable polynomials in $L[x]$.

DEFINITION 10.7.10. The *absolute Galois group* of a field K is the Galois group

$$G_K = \mathrm{Gal}(K^{\mathrm{sep}}/K),$$

where K^{sep} is a separable closure of K.

REMARK 10.7.11. The absolute Galois group, despite the word "the", is not unique, but rather depends on the choice of separable closure. An isomorphism of separable closures gives rise to a canonical isomorphism of absolute Galois groups, however.

EXAMPLE 10.7.12. Let q be a power of a prime number. Then there is a unique topological isomorphism $G_{\mathbb{F}_q} \xrightarrow{\sim} \hat{\mathbb{Z}}$ sending the Frobenius automorphism $\varphi_q \colon x \mapsto x^q$ to 1. To see this, note that $\mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q) \to \mathbb{Z}/n\mathbb{Z}$ given by sending φ_q to 1 is an isomorphism, and these give rise to compatible isomorphisms in the inverse limit

$$G_{\mathbb{F}_q} \xrightarrow{\sim} \varprojlim_n \mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q) \xrightarrow{\sim} \varprojlim_n \mathbb{Z}/n\mathbb{Z} \xrightarrow{\sim} \hat{\mathbb{Z}}.$$

EXAMPLE 10.7.13. Let $\mathbb{Q}(\mu_{p^\infty})$ denote the field given by adjoining all p-power roots of unity to $\mathbb{Q}$. Then

$$\mathrm{Gal}(\mathbb{Q}(\mu_{p^\infty})/\mathbb{Q}) \cong \varprojlim_n \mathrm{Gal}(\mathbb{Q}(\mu_{p^n})/\mathbb{Q}) \cong \varprojlim_n (\mathbb{Z}/p^n\mathbb{Z})^\times \cong \mathbb{Z}_p^\times$$

the middle isomorphisms arising from the p^nth cyclotomic characters.

TERMINOLOGY 10.7.14. The isomorphism $\mathrm{Gal}(\mathbb{Q}(\mu_{p^\infty})/\mathbb{Q}) \to \mathbb{Z}_p^\times$ of Example 10.7.13 called the *p-adic cyclotomic character*.

Since the compositum of two abelian extensions of a field inside a fixed algebraic closure is abelian, the following makes sense.

NOTATION 10.7.15. Let K be a field. The *maximal abelian extension* of K inside an algebraic closure of K is denoted K^{ab}.

REMARK 10.7.16. The abelianization G_K^{ab} of the absolute Galois group G_K of a field K canonically isomorphic to $\mathrm{Gal}(K^{\mathrm{ab}}/K)$ via the map induced by restriction on G_K.

CHAPTER 11

Commutative algebra

In this chapter, all rings are commutative with unity.

11.1. Localization

We have previously discussed quotient fields, or fields of fractions, of integral domains. In this section, we generalize the notion to arbitrary commutative rings with unity and allow subrings with denominators in a smaller set.

DEFINITION 11.1.1. A subset S of R is *multiplicatively closed* if it is closed under multiplication, $1 \in S$, and $0 \notin S$.

We now begin to generalize our earlier constructions. First, we prove a strengthening of Lemma 3.7.1.

LEMMA 11.1.2. *Let R be a commutative ring, and let S be a multiplicatively closed subset of R. The relation $\sim$ on $R \times S$ given by $(a,s) \sim (b,t)$ if and only if there exists $r \in S$ such that $rat = rbs$ is an equivalence relation.*

PROOF. Let $a,b,c \in R$ and $s,t,u \in S$. That $\sim$ is reflexive is the fact that $ras = ras$ for any $r \in S$, that it is symmetric is the fact that $rat = rbs$ implies $rbs = rat$. If $q,r \in S$ are such that $rat = rbs$ and $qbu = qct$, then multiplying the former equality by qu and then applying the latter, we obtain

$$(rqt)au = q(rat)u = q(rbs)u = r(qbu)s = r(qct)s = (rqt)cs.$$

We have $rqt \in S$ since S is multiplicatively closed, so $(a,s) \sim (c,u)$. Therefore, $\sim$ is transitive. □

REMARK 11.1.3. Let R be a commutative ring, and let S be a multiplicatively closed subset of R. If S contains no zero divisors, then the relation $\sim$ on $R \times S$ is more simply defined by $(a,s) \sim (b,t)$ if and only if $at = bs$. That is, this implies $rat = rbs$ for all $r \in R$, and likewise, the latter implies $at = bs$ since r is not a zero divisor.

DEFINITION 11.1.4. Let R be a commutative ring and S a multiplicatively closed subset of R. The equivalence class $\frac{a}{s}$ of a pair $(a,s) \in R \times S$ is called an fraction of R with denominator in S (or S-fraction), and the set of such S-fractions is denoted $S^{-1}R$.

REMARK 11.1.5. Let R be a commutative ring and S a multiplicatively closed subset of R. By definition, we have $\frac{a}{s} = \frac{at}{st}$ for any $a \in R$ and $s,t \in S$. We denote the fraction $\frac{a}{1}$ more simply by a.

REMARK 11.1.6. If we were to allow $0 \in S$, then $S^{-1}R$ would have just one element 0. The condition that $1 \in S$ is not strictly necessary so long as S is nonempty, as we can set $a = \frac{as}{s}$ for any $s \in S$ anyway.

We leave the proof of the following to the reader.

THEOREM 11.1.7. *The set $S^{-1}R$ is a ring under addition and multiplication of fractions:*

$$\frac{a}{s} + \frac{b}{t} = \frac{at+bs}{st} \quad \text{and} \quad \frac{a}{s} \cdot \frac{b}{t} = \frac{ab}{st}.$$

PROOF. Suppose that $(a,s) \sim (a',s')$, and let $r \in R$ be such that $ras' = ra's$. Then

$$r(at+bs)s't = ras'tt + rbss't = ra'stt + rbss't = r(a't+bs')st,$$

so addition is well-defined, noting its symmetry. Similarly, we have

$$rabs't = ra'bst,$$

so multiplication is well-defined. By definition, addition and multiplication are commutative, and associativity and distributivity of the two are exactly as in the proof of Theorem 3.7.6. Moreover, for any $s,t \in S$, we have

$$\frac{a}{s} + 0 = \frac{a}{s} + \frac{0}{t} = \frac{at}{st} = \frac{a}{s} \quad \text{and} \quad \frac{a}{s} \cdot 1 = \frac{a}{s} \cdot \frac{t}{t} = \frac{at}{st} = \frac{a}{s}.$$

Also, we have

$$\frac{a}{s} + \frac{-a}{s} = \frac{as + (-a)s}{s^2} = 0,$$

so $-\frac{a}{s} = \frac{-a}{s}$. Thus, $S^{-1}R$ is a ring under addition and multiplication. □

REMARK 11.1.8. Theorem 11.1.10 tells us that if S has no zero divisors, then $S^{-1}R$ is the smallest ring containing R in which every element of S is a unit. In particular, if every element of S already is a unit in R, then $S^{-1}R = R$. If S has zero divisors, then the map from S to the unit group of $S^{-1}R$ is still injective.

DEFINITION 11.1.9. The ring $S^{-1}R$ consisting of S-fractions for a multiplicatively closed subset S of a commutative ring R with unity is called the ring of S-fractions of R, or the localization of R at S.

THEOREM 11.1.10. *Let R be a ring and S a multiplicatively closed subset of R.*

a. There is a canonical ring homomorphism $R \to S^{-1}R$ given by $\phi_S(a) = a$ for all $a \in R$, and it is injective if and only if S contains no zero divisors in R.

b. Every element of S maps to a unit in $S^{-1}R$ under ϕ_S.

c. If Q is a commutative ring and $f\colon R \to Q$ is a homomorphism such that $f(S) \subset Q^{\times}$, then there is a unique injective homomorphism $\theta\colon S^{-1}R \to Q$ such that $f = \theta \circ \phi_S$.

PROOF. That ϕ_S is a homomorphism is simply that

$$\phi_S(a)\phi_S(b) = a \cdot b = \frac{as}{s} \cdot \frac{bt}{t} = \frac{abst}{st} = ab = \phi_S(ab)$$

for any $s,t \in S$. If S contains no zero divisors $\phi_S(a) = 0$ for some $a \in R$, then $(as,s) \sim (0,t)$, so $ast = 0$, which implies that $a = 0$. Similarly, if ϕ_S is injective, then $ast \neq 0$ for all nonzero $a \in R$ and elements $s,t \in S$, which means that $as \neq 0$ for all nonzero $a \in R$ and elements $s \in S$.

Note that $\frac{1}{s} \in S^{-1}R$ is clearly a multiplicative inverse of $\phi_S(s) \in S^{-1}R$. Define $\theta \colon S^{-1}R \to Q$ by $\theta(\frac{a}{s}) = f(s)^{-1}f(a)$. It is easily checked to be a homomorphism. It also restricts to f by definition. To see that it is well-defined, note that if $\frac{a}{s} = \frac{b}{t}$, then $at = bs$, and so $s^{-1}a = t^{-1}b$ in $S^{-1}R$ and thus $f(s)^{-1}f(a) = f(t)^{-1}f(b)$ in Q. □

NOTATION 11.1.11. For an ideal I of a commutative ring R and a multiplicative set S in R, let $S^{-1}I$ denote the ideal generated by the image $\phi_S(I)$ of I in $S^{-1}R$.

PROPOSITION 11.1.12. *Let R be a commutative ring, and let S be a multiplicative subset of R.*

a. *For any ideal I of R, we have*
$$\phi_S^{-1}(S^{-1}I) = \{a \in R \mid Sa \cap I \neq \varnothing\}.$$

b. *For any ideal J of $S^{-1}R$, we have $S^{-1}\phi_S^{-1}(J) = J$.*

PROOF. First, we remark that every element of $S^{-1}I$ has the form $\frac{a}{s}$ with $a \in I$ and $s \in S$. That is, by definition, every element of $S^{-1}I$ is an $S^{-1}R$-linear combination of fractions $\frac{a}{1}$ with $a \in I$, so an R-linear combination of fractions $\frac{a}{s}$ with $a \in I$ and $s \in S$. But we can take common denominators and use the fact that I is an R-ideal to write every such fraction in the desired form.

Let $a \in R$ be such that we have $s \in S$ with $x = sa \in I$. Then $\phi_S(a) = \frac{x}{s} \in S^{-1}I$, so $a \in \phi_S^{-1}(S^{-1}I)$. Conversely, if $\phi_S(a) = \frac{x}{s}$ for some $x \in I$ and $s \in S$, then $rsa = rx$ for some $r \in S$, from which it follows that $Sa \cap I \neq \varnothing$. This proves part a.

For part b, take $\frac{x}{s} \in J$ with $x \in R$ and $s \in S$, and note that $x \in \phi_S^{-1}(J)$. We then have that $\frac{x}{s} \in S^{-1}\phi_S^{-1}(J)$ by definition. On the other hand, the image in J of any element of $\phi_S^{-1}J$ is clearly in J by definition as well. Thus, we have part b. □

DEFINITION 11.1.13. Let R be a commutative ring, and let S be a multiplicative subset of R.

a. For any ideal I of R, the *expansion* of I in $S^{-1}R$ is $S^{-1}I$.

b. For any ideal J of $S^{-1}R$, the *contraction* of J in R is $\phi_S^{-1}(J)$.

PROPOSITION 11.1.14. *Let R be a commutative ring, and let S be a multiplicatively closed subset of R. Then contraction and expansion given mutually inverse maps between the set of prime ideals of R disjoint from S and the set of prime ideals of $S^{-1}R$. Moreover, the expansion of any prime ideal of R that intersects S is $S^{-1}R$.*

PROOF. By Proposition 11.1.12, we need merely note that for a prime ideal $\mathfrak{p}$ of R, the ideal
$$J = \phi_S^{-1}(S^{-1}\mathfrak{p}) = \{a \in R \mid Sa \cap \mathfrak{p} \neq \varnothing\}$$
is $\mathfrak{p}$ if $\mathfrak{p} \cap S = \varnothing$ and $S^{-1}R$ otherwise. The ideal J clearly contains $\mathfrak{p}$. If $a \in J - \mathfrak{p}$, then there exists $s \in S$ such that $sa \in \mathfrak{p}$. But then $s \in \mathfrak{p}$ by the primality of $\mathfrak{p}$, so $\mathfrak{p}$ and S are not disjoint. Moreover, in this case, $S^{-1}\mathfrak{p}$ contains $1 = \frac{s}{s}$, hence $S^{-1}\mathfrak{p} = R$, and therefore $J = S^{-1}R$. □

DEFINITION 11.1.15. The *total ring of fractions* $Q(R)$ of a commutative ring R with unity is the localization of R at the set of nonzero elements of R that are not zero divisors.

EXAMPLES 11.1.16.

a. If R is an integral domain, then its total ring of fractions is its field of fractions $Q(R)$.

b. If $R = \mathbb{Z} \times \mathbb{Z}$, then its total ring of fractions is given by inverting the set

$$S = \{(c,d) \mid c,d \in \mathbb{Z} - \{0\}\}.$$

There is a ring isomorphism

$$Q(\mathbb{Z}\times\mathbb{Z}) \xrightarrow{\sim} \mathbb{Q}\times\mathbb{Q}, \qquad \tfrac{(a,b)}{(c,d)} \mapsto (\tfrac{a}{c}, \tfrac{b}{d}).$$

DEFINITION 11.1.17. Let R be a commutative ring, and let $x \in R$. Then the *localization of R with respect to x*, denoted by R_x and also by $R[x^{-1}]$, is the ring $S^{-1}R$ for $S = \{x^n \mid n \geq 0\}$.

EXAMPLE 11.1.18. Let $n \in \mathbb{Z}$. Then the ring $\mathbb{Z}[\frac{1}{n}]$ may be identified with the subset of $\mathbb{Q}$ consisting of reduced fractions with denominator a product of powers of primes dividing n, or equivalently, with denominator dividing a power of n. The distinct ideals of $\mathbb{Z}[\frac{1}{n}]$ are generated by nonnegative $a \in \mathbb{Z}$ with $(a,n) = 1$.

EXAMPLE 11.1.19. Let $R = \mathbb{Z}\times\mathbb{Z}$ and $x = (1,0)$, and consider R_x. Since $x \cdot (0,1) = 0$, and x is invertible in R_x, we have $(0,1) = 0$ in R_x. Note also that $(1,0)^n(a,0) = (a,0)$, so $\frac{(a,0)}{(1,0)^n} = (a,0)$. It follows that the ring homomorphism $\mathbb{Z} \to R_x$ given by $a \mapsto (a,0)$ is an isomorphism.

LEMMA 11.1.20. *Let R be a commutative ring, and let $\mathfrak{p}$ be a prime ideal of R. Then $R - \mathfrak{p}$ is a multiplicatively closed subset of R.*

PROOF. If $a,b \in S$, then since $a,b \notin \mathfrak{p}$ and $\mathfrak{p}$ is prime, we have $ab \notin \mathfrak{p}$, so $ab \in S$. Moreover $0 \notin S$ and $1 \in S$ by definition. □

DEFINITION 11.1.21. Let R be a commutative ring, and let $\mathfrak{p}$ be a prime ideal of R. Then $R_{\mathfrak{p}} = S_{\mathfrak{p}}^{-1}R$, where $S_{\mathfrak{p}}^{-1} = R - \mathfrak{p}$, is referred to as *the localization of R at* $\mathfrak{p}$.

EXAMPLES 11.1.22.

a. Let $p \in \mathbb{Z}$ be prime. Then $\mathbb{Z}_{(p)}$ may be identified with the subring of $\mathbb{Q}$ consisting of reduced fractions with denominators not divisible by p.

b. The ring $\mathbb{Q}[x]_{(x)}$ consists of rational functions with denominator not divisible by x. Contrast this with $\mathbb{Q}[x]_x = \mathbb{Q}[x,x^{-1}]$, which consists of rational functions with denominator a power of x.

EXAMPLES 11.1.23.

a. The ring $\mathbb{Z}[x]_{(x)}$ is identified with $\mathbb{Q}[x]_{(x)}$ inside $\mathbb{Q}(x)$.

b. The ring $\mathbb{Z}[x]_{(p,x)}$ is the subring of $\mathbb{Q}(x)$ of rational functions with denominator having nonzero constant term modulo p.

LEMMA 11.1.24. *Let M be a module over a commutative ring R, and let S be a multiplicatively closed subset of R. Then the relation $\sim_S$ on $S \times M$ defined by $(s,m) \sim (t,n)$ if there exists $r \in S$ such that $r(sn - tm) = 0$ is an equivalence relation.*

PROOF. The relation $\sim$ is clearly reflexive and symmetric, so we only need to check transitivity. For this, let $(s,m) \sim (s',m')$ and $(s',m') \sim (s'',m'')$ in $S \times M$. Then there exist $r, r' \in S$ such that $r(sm' - s'm) = r'(s'm'' - s''m') = 0$. We then have

$$0 = rr's''(sm' - s'm) + rr's(s'm'' - s''m') = rr's'(sm'' - s''m),$$

so $(s,m) \sim (s'',m'')$. □

NOTATION 11.1.25. Let M be a module over a commutative ring R, and let S be a multiplicatively closed subset of R. The set of equivalence classes of $S \times M$ under $\sim_S$ is denoted $S^{-1}M$, and the equivalence class of (s,m) is denoted $s^{-1}m$ or $\frac{m}{s}$. We write $\frac{m}{1}$ more simply as m.

We omit the easy but nonetheless tedious proof of the following.

PROPOSITION 11.1.26. *Let M be a module over a commutative ring R, and let S be a multiplicatively closed subset of R. The set $S^{-1}M$ of equivalence classes of $S \times M$ under the equivalence relation $\sim_S$ is an $S^{-1}R$-module under the operations*

$$\frac{m}{s} + \frac{n}{t} = \frac{tm + ns}{st} \quad \textit{and} \quad \frac{a}{s} \cdot \frac{m}{t} = \frac{am}{st}$$

for $a \in R$, $m, n \in M$, and $s, t \in S$. There is a canonical map $\iota \colon M \to S^{-1}M$ of R-modules given by $\iota(m) = \frac{m}{1}$.

EXAMPLE 11.1.27. Let S be a multiplicatively closed subset of a commutative ring R. Then the localization $S^{-1}R$ of R viewed as a left R-module is just the ring $S^{-1}R$ viewed as a module over itself.

EXAMPLE 11.1.28. Let R be an integral domain and $S = R - \{0\}$. If M is an R-module, then $S^{-1}M$ is a $Q(R)$-vector space.

LEMMA 11.1.29. *Let S be a multiplicatively closed subset of a commutative ring R. Let $\{M_i \mid i \in I\}$ be a collection of R-modules. Then*

$$S^{-1}\Big(\bigoplus_{i \in I} M_i\Big) \xrightarrow{\sim} \bigoplus_{i \in I} S^{-1}M_i$$

via the canonical map that takes $s^{-1}(m_i)_{i \in I}$ to $(s^{-1}m_i)_{i \in I}$.

EXAMPLE 11.1.30. Let p be a prime number, and let S_p be the multiplicatively closed subset of $\mathbb{Z}$ that is the complement of the prime ideal (p). For $n \geq 1$, the localization $S_p^{-1}(\mathbb{Z}/n\mathbb{Z})$ is isomorphic to $\mathbb{Z}/p^k\mathbb{Z}$, where p^k is the highest power of p dividing n.

To see this, note that $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/p^k\mathbb{Z}$, where $n = p^k m$, so we have

$$S_p^{-1}(\mathbb{Z}/n\mathbb{Z}) = S_p^{-1}(\mathbb{Z}/m\mathbb{Z}) \times S_p^{-1}(\mathbb{Z}/p^k\mathbb{Z}).$$

Now, for any $x \in \mathbb{Z}/m\mathbb{Z}$, we have $x = m^{-1}(mx) = m^{-1}0 = 0$ in $S_p^{-1}(\mathbb{Z}/m\mathbb{Z})$. It follows that $S_p^{-1}(\mathbb{Z}/m\mathbb{Z}) = 0$. On the other hand, if $y \in \mathbb{Z}/p^k\mathbb{Z}$ and $a \in S_p$ are such that $a^{-1}y = 0$, then there exists $b \in S_p$ such that $by = 0$, which means that $y = 0$ since b is prime to p. Furthermore, any $a^{-1}y \in S_p^{-1}(\mathbb{Z}/p^k\mathbb{Z})$ is in the image of $\mathbb{Z}/p^k\mathbb{Z}$ as $a \in (\mathbb{Z}/p^k\mathbb{Z})^\times$. Thus, we have $S_p^{-1}(\mathbb{Z}/p^k\mathbb{Z}) = \mathbb{Z}/p^k\mathbb{Z}$.

PROPOSITION 11.1.31. *Let M be a module over a commutative ring R, and let S be a multiplicatively closed subset of R. Then $S^{-1}M \cong S^{-1}R \otimes_R M$ as $S^{-1}R$-modules.*

PROOF. Define a map $\theta\colon S^{-1}R \times M \to S^{-1}M$ by $\theta(s^{-1}a, m) = \frac{am}{s}$. To see that it is well-defined, note that if $\frac{a}{s} = \frac{b}{t}$, then we have $r \in S$ with $r(ta - sb) = 0$, and then $r(tam - sbm) = 0$, so $\frac{am}{s} = \frac{bm}{t}$. The map θ is easily checked to be left $S^{-1}R$-linear, right R-linear, and R-balanced. We then obtain a map of $S^{-1}R$-modules $\Theta\colon S^{-1}R \otimes_R M \to S^{-1}M$ satisfying $\Theta(\frac{a}{s} \otimes m) = \frac{am}{s}$ by the universal property of the tensor product. (That it is an $S^{-1}R$-module homomorphism, rather than just an R-module homomorphism, follows directly from the left $S^{-1}R$-linearity.) For bijectivity, it suffices to exhibit an inverse function.

Define a function $\psi\colon S \times M \to S^{-1}R \otimes_R M$ by $\psi(s,m) = s^{-1} \otimes m$. If $(s,m) \sim_S (t,n)$, then let $r \in S$ be such that $r(sn - tm) = 0$. We then have

$$s^{-1} \otimes m = (rst)^{-1} \otimes rtm = (rst)^{-1} \otimes rsn = t^{-1} \otimes n,$$

so we obtain a well-defined map $\Psi\colon S^{-1}M \to S^{-1}R \otimes M$ given by $\Psi(\frac{m}{s}) = s^{-1} \otimes m$. (In fact, Ψ is a homomorphism of $S^{-1}R$-modules, but it is not necessary to check this to finish the proof, since the inverse of a module isomorphism is one as well.) By definition, $\Theta(\Psi(\frac{m}{s})) = \frac{m}{s}$, and we have

$$\Psi(\Theta(s^{-1}a \otimes m)) = \Psi(\tfrac{am}{s}) = s^{-1} \otimes am = s^{-1}a \otimes m.$$

□

REMARK 11.1.32. Given a commutative ring R and a multiplicatively closed set S, localization provides a functor $S^{-1}\colon R\textbf{-mod} \to S^{-1}R\textbf{-mod}$. That is, if $f\colon M \to N$ is an R-modules homomorphism, then we have an induced R-module homomorphism $S^{-1}f\colon S^{-1}M \to S^{-1}N$ given by $f(s^{-1}m) = s^{-1}f(m)$.

11.2. Local rings

DEFINITION 11.2.1. A commutative ring R is *local* if it has a unique maximal ideal.

DEFINITION 11.2.2. The *residue field* of a local ring R with maximal ideal $\mathfrak{m}$ is the field $R/\mathfrak{m}$.

The first part of the following explains something of the meaning of the terminology "localization."

PROPOSITION 11.2.3. *Let $\mathfrak{p}$ be a prime ideal of a commutative ring R.*

a. The ring $R_\mathfrak{p}$ is a local ring with maximal ideal $\mathfrak{p}R_\mathfrak{p}$.

b. The proper ideals of $R_\mathfrak{p}$ are exactly those of the form $IR_\mathfrak{p}$ for some ideal I of R contained in $\mathfrak{p}$.

PROOF. By Proposition 11.1.12b, every ideal of $R_{\mathfrak{p}}$ has the form $IR_{\mathfrak{p}}$ for some ideal I of R. If $a \in R - \mathfrak{p}$, then by definition a is invertible in $R_{\mathfrak{p}}$, hence $aR_{\mathfrak{p}} = R_{\mathfrak{p}}$. Thus, $IR_{\mathfrak{p}} = R_{\mathfrak{p}}$ for every ideal of R not contained in $\mathfrak{p}$. On the other hand $IR_{\mathfrak{p}} \subseteq \mathfrak{p}R_{\mathfrak{p}}$ if $I \subseteq \mathfrak{p}$, so $\mathfrak{p}R_{\mathfrak{p}}$ is the unique maximal ideal of $R_{\mathfrak{p}}$. □

We note the following easy lemmas.

LEMMA 11.2.4. *Let R be a local ring and $\mathfrak{m}$ be its maximal ideal. Then $R^{\times} = R - \mathfrak{m}$.*

PROOF. If $(a) \neq R$, then a is contained in a maximal ideal, which must be $\mathfrak{m}$. Conversely, if $a \in R^{\times}$, then $(a) = R$, so a is not contained in $\mathfrak{m}$. □

LEMMA 11.2.5. *Let $\mathfrak{m}$ be a maximal ideal of a commutative ring R. Then the canonical ring homomorphism $R/\mathfrak{m} \to R_{\mathfrak{m}}/\mathfrak{m}R_{\mathfrak{m}}$ is an isomorphism.*

PROOF. Since nonzero maps of fields are injective, it suffices to see that the map is onto. If $r \in R$ and $u \in R - \mathfrak{m}$, then let $v \in R$ be such that $(u + \mathfrak{m})(v + \mathfrak{m}) = 1$. Then $\frac{r}{u} + \mathfrak{m}R_{\mathfrak{m}}$ is the image of $vr + \mathfrak{m}$. □

NOTATION 11.2.6. Let $\mathfrak{p}$ be a prime ideal of a commutative ring R, and let M be an R-module. Then the localization of the $R_{\mathfrak{p}}$-module $S_{\mathfrak{p}}^{-1}M$ is denoted $M_{\mathfrak{p}}$.

PROPOSITION 11.2.7. *Let R be a commutative ring and M be an R-module. Then the following are equivalent:*

i. $M = 0$,

ii. $M_{\mathfrak{p}} = 0$ for every prime ideal $\mathfrak{p}$ of R, and

iii. $M_{\mathfrak{m}} = 0$ for every maximal ideal $\mathfrak{m}$ of R.

PROOF. Clearly, (i) implies (ii) and (ii) implies (iii). Let $m \in M$ be nonzero. Let I be annihilator of m in R, which is to say $I = \mathrm{Ann}(Rm)$. Then I is a proper ideal, hence contained in a maximal ideal $\mathfrak{m}$ of R. If $\frac{r}{u} \in R_{\mathfrak{m}}$ annihilates m, then $rsm = 0$ for some $s \in R - \mathfrak{m}$. Thus $rs \in \mathfrak{m}$, so $r \in \mathfrak{m}$ as $\mathfrak{m}$ is prime. This implies that the annihilator of m in $R_{\mathfrak{m}}$ in a proper ideal, so m is nonzero in $M_{\mathfrak{m}}$. Thus, we have the contrapositive to (iii) implies (i). □

DEFINITION 11.2.8. The *Jacobson radical* $J(R)$ of a ring R is the intersection of all left maximal ideals of R.

The following extends Lemma 11.2.4.

LEMMA 11.2.9. *Let $x \in R$. Then $x \in J(R)$ if and only if $1 - rx \in R^{\times}$ for all $r \in R$.*

PROOF. If $1 - rx \notin R^{\times}$, then there exists a left maximal ideal $\mathfrak{m}$ containing $1 - rx$. Then $rx \notin \mathfrak{m}$, so $rx \notin J(R)$, and therefore $x \notin J(R)$. Conversely, if $x \notin J(R)$, then there exists a left maximal ideal $\mathfrak{m}$ such that $x \notin \mathfrak{m}$. Then there exist $r \in R$ and $y \in \mathfrak{m}$ such that $1 = rx + y$. Then $1 - rx = y \notin R^{\times}$. □

THEOREM 11.2.10 (Nakayama's lemma). *Let M be a finitely generated module over a commutative ring R, and suppose that $J(R)M = M$. Then $M = 0$.*

PROOF. Let $\{m_1, m_2, \dots, m_k\}$ be a set of generators of M with $k \geq 1$. Since $m_1 \in J(R)M$, we can find $a_i \in J(R)$ for $1 \leq i \leq k$ such that

$$m_1 = \sum_{i=1}^{k} a_i m_i.$$

Since $(1 - a_1)m_1$ is contained in the submodule M' generated by $m_2, \dots, m_k$. On the other hand, $1 - a_1 \in R^\times$ by Lemma 11.2.9. But then m_1 itself is contained in M', which tells us that $M' = M$ and k is not minimal. That is, the minimal number of generators of M is zero. □

COROLLARY 11.2.11. *Let M be a finitely generated module over a local ring R, and suppose that $\mathfrak{m}M = M$, where $\mathfrak{m}$ is the maximal ideal of M. Then $M = 0$.*

COROLLARY 11.2.12. *Let M be a finitely generated module over a local ring R with maximal ideal $\mathfrak{m}$, and let X be a set of elements of M such that $\{m + \mathfrak{m}M \mid m \in X\}$ generates $M/\mathfrak{m}M$ as a vector space over the residue field $R/\mathfrak{m}$. Then X generates M.*

PROOF. Let N be the submodule of M generated by X. Then $N + \mathfrak{m}M = M$, so every element in M/N is the N-coset of some element of $\mathfrak{m}M$, which is to say that $\mathfrak{m}(M/N) = M/N$. By Nakayama's lemma, we have $M/N = 0$, so X generates M. □

EXAMPLE 11.2.13. Take the set of tuples $(111, 107, 50)$, $(23, -17, 41)$, and $(30, -8, 104)$. Suppose that we want to see if they generate the $\mathbb{Q}$-vector space $\mathbb{Q}^3$. It suffices, then, to see that they generate the $\mathbb{Z}_{(p)}$-module $\mathbb{Z}_{(p)}$ for some prime p. Moreover, the map $\mathbb{F}_p \to \mathbb{Z}_{(p)}/p\mathbb{Z}_{(p)}$ is an isomorphism, so by Corollary 11.2.12, it suffices to see that these tuples generate $\mathbb{F}_p^3$. Modulo 2, they are $(1,1,0)$, $(1,1,1)$, and $(0,0,0)$, so they do not generate $\mathbb{F}_2^3$. However, modulo 3, they are $(0,-1,-1)$, $(-1,1,-1)$, and $(0,1,-1)$, which do in fact generate $\mathbb{F}_3^3$, and thus the original tuples generate $\mathbb{Q}^3$.

Nakayama's lemma can also be used to prove the following result for free modules.

LEMMA 11.2.14. *Let M be a finitely generated free module over a local ring R with maximal ideal $\mathfrak{m}$, and let X be a subset of M. If the image of X in $M/\mathfrak{m}M$ is R-linearly independent, then X is R-linearly independent and can be extended to a basis of M.*

PROOF. Let $\bar{X}$ denote the image of X in $M/\mathfrak{m}M$. Extend $\bar{X}$ to a basis $\bar{B}$ of $M/\mathfrak{m}M$, and let $B \subset M$ be a lift of $\bar{B}$ to M with $X \subseteq B$. Then B spans M by Corollary 11.2.12. To see that it is linearly independent, suppose that B has n elements $m_1, \dots, m_n$ and consider the sum $\sum_{i=1}^n a_i m_i$ for some $a_i \in R$. Suppose that not all a_i are zero, and let $k \geq 0$ be minimal such that $a_i \in \mathfrak{m}^k$ for all i. Note that the map

$$\mathfrak{m}^k/\mathfrak{m}^{k+1} \otimes_R M \to \mathfrak{m}^k M/\mathfrak{m}^{k+1} M$$

induced by the R-action on M is an isomorphism by the freeness of M, since tensor products commute with direct sums and it is clearly true for $M = R$. But we have $\sum_{i=1}^n a_i \otimes m_i \neq 0$ in the left-hand side (which is isomorphic to $\mathfrak{m}^k/\mathfrak{m}^{k+1} \otimes_{R/\mathfrak{m}} M/\mathfrak{m}M$) since $\bar{B}$ is a basis of $M/\mathfrak{m}M$. Therefore $\sum_{i=1}^n a_i m_i \neq 0$. In other words B is a basis of M, and X is R-linearly independent. □

11.3. Integral extensions

DEFINITION 11.3.1. We say that B/A is an *extension of commutative rings* if A and B are commutative rings such that A is a subring of B.

DEFINITION 11.3.2. Let B/A be an extension of commutative rings. We say that $\beta \in B$ is *integral over A* if β is the root of a monic polynomial in $A[x]$.

EXAMPLES 11.3.3.

a. Every element $a \in A$ is integral over A, in that a is a root of $x-a$.

b. If L/K is a field extension and $\alpha \in L$ is algebraic over K, then α is integral over K, being a root of its minimal polynomial, which is monic.

c. If L/K is a field extension and $\alpha \in L$ is transcendental over K, then α is not integral over K.

d. The element $\sqrt{2}$ of $\mathbb{Q}(\sqrt{2})$ is integral over $\mathbb{Z}$, as it is a root of x^2-2.

e. The element $\alpha = \frac{1-\sqrt{5}}{2}$ of $\mathbb{Q}(\sqrt{5})$ is integral over $\mathbb{Z}$, as it is a root of x^2-x-1.

PROPOSITION 11.3.4. *Let B/A be an extension of commutative rings. For $\beta \in B$, the following conditions are equivalent:*

i. the element β is integral over A,

ii. there exists $n \geq 0$ such that $\{1,\beta,\dots,\beta^n\}$ generates $A[\beta]$ as an A-module,

iii. the ring $A[\beta]$ is a finitely generated A-module, and

iv. there exists a faithful $A[\beta]$-submodule M of B that is finitely generated over A.

PROOF. Suppose that (i) holds. Then β is a root of a monic polynomial $g \in A[x]$. Given any $f \in A[x]$, the division algorithm tells us that $f = qg + r$ with $q, r \in A[x]$ and either $r = 0$ or $\deg r < \deg g$. It follows that $f(\beta) = r(\beta)$, and therefore that $f(\beta)$ is in the A-submodule generated by $\{1,\beta,\dots,\beta^{\deg g-1}\}$, so (ii) holds. Since this set is independent of f, it generates $A[\beta]$ as an A-module, so (iii) holds. Suppose that (iii) holds. Then we may take $M = A[\beta]$, which being free over itself has trivial annihilator.

Finally, suppose that (iv) holds. Let

$$M = \sum_{i=1}^{n} A\gamma_i \subseteq B$$

be such that $\beta M \subseteq M$, and suppose without loss of generality that $\beta \neq 0$. We have

$$\beta\gamma_j = \sum_{j=1}^{n} a_{ij}\gamma_i$$

for some $a_{ij} \in A$ with $1 \leq i, j \leq n$. Consider A-module homomorphism $T \colon B^n \to B^n$ represented by (a_{ij}). The characteristic polynomial $c_T(x) \in A[x]$ is monic, and $c_T(\beta)$ acts as zero on M. Since M is a faithful $A[\beta]$-module, we must have $c_T(\beta) = 0$. Thus, β is integral. □

EXAMPLE 11.3.5. The element $\frac{1}{2} \in \mathbb{Q}$ is not integral over $\mathbb{Z}$, as $\mathbb{Z}[1, 2^{-1}, \dots, 2^{-n}]$ for $n \geq 0$ is equal to $\mathbb{Z}[2^{-n}]$, which does not contain $2^{-(n+1)}$.

DEFINITION 11.3.6. Let B/A be an extension of commutative rings. We say that B is an *integral extension of A* if every element of B is integral over A.

EXAMPLE 11.3.7. The ring $\mathbb{Z}[\sqrt{2}]$ is an integral extension of $\mathbb{Z}$. Given $\alpha = a + b\sqrt{2}$ with $a, b \in \mathbb{Z}$, note that α is a root of $x^2 - 2ax + a^2 - 2b^2$.

The integral extensions of a field are its algebraic field extensions.

LEMMA 11.3.8. *Let B be a domain that is integral over a field F. Then B is a field that is an algebraic extension of F.*

PROOF. Every $b \in B \subseteq Q(B)$ is the root of a polynomial with coefficients in F, so $F(b) = F[b] \subseteq B$. That is, $b \in B^\times$, and thus B is a field and algebraic over F. □

LEMMA 11.3.9. *Suppose that B/A is an extension of commutative rings such that B is finitely generated as an A-module, and let M be a finitely generated B-module. Then M is a finitely generated A-module.*

PROOF. Let $\{m_1, \dots, m_n\}$ be a set of generators of M as a B-module, and let $\{\beta_1, \dots, \beta_k\}$ be a set of generators of B as an A-module. We claim that $\{\beta_i m_j \mid 1 \leq i \leq k, 1 \leq j \leq n\}$ is a set of generators of M as an A-module. To see this, let $m \in M$ and write

$$m = \sum_{j=1}^{n} b_j m_j$$

with $b_j \in B$ for $1 \leq j \leq n$. For $1 \leq j \leq n$, we then write

$$b_j = \sum_{i=1}^{k} a_{ij} \beta_i$$

with $a_{ij} \in A$ for $1 \leq i \leq k$. We then have

$$m = \sum_{i=1}^{k} \sum_{j=1}^{n} a_{ij} \beta_i m_j,$$

as desired. □

We now give a criterion for a finitely generated algebra over a ring to be finitely generated as a module.

PROPOSITION 11.3.10. *Let B/A be an extension of commutative rings and suppose that*

$$B = A[\beta_1, \beta_2, \dots, \beta_k]$$

for some $k \geq 0$ and $\beta_i \in B$ with $1 \leq i \leq k$. Then the following are equivalent:

i. the ring B is integral over A,

ii. each β_i with $1 \leq i \leq k$ is integral over A, and *iii. the ring B is finitely generated as an A-module.*

PROOF. Clearly, (i) implies (ii), so suppose that (ii) holds. By definition, each β_i is then integral over any commutative ring containing A. By Proposition 11.3.4, each $A[\beta_1,\dots,\beta_j]$ with $1 \le j \le k$ is a finitely generated $A[\beta_1,\dots,\beta_{j-1}]$-module, generated by $\{1,\beta_j,\dots,\beta_j^{n_j}\}$ for some $n_j \ge 0$. Assuming recursively that $A[\beta_1,\dots,\beta_{j-1}]$ is finitely generated as an A-module, Lemma 11.3.9 implies that $A[\beta_1,\dots,\beta_j] = A[\beta_1,\dots,\beta_{j-1}][\beta_j]$ is finitely generated as an A-module as well. Therefore, (iii) holds. Finally, if (iii) holds and $\beta \in B$, then since $\beta B \subseteq B$, the element β is integral over a by Proposition 11.3.4. Thus (i) holds. □

We derive the following important consequence.

PROPOSITION 11.3.11. *Suppose that C/B and B/A are integral extensions of commutative rings. Then C/A is an integral extension as well.*

PROOF. Let $\gamma \in C$, and let $f \in B[x]$ be a monic polynomial which has γ as a root. Let B' be the subring of B generated over A by the coefficients of f, which is integral over A as B is. By Proposition 11.3.10, the ring B' is then finitely generated over A. As $B'[\gamma]$ is finitely generated over B' as well, we have $B'[\gamma]$ is finitely generated over A. Hence, $B[\gamma]$ is itself an integral extension of A. By definition of an integral extension, the element γ is integral over A. Since $\gamma \in C$ was arbitrary, we conclude that C is integral over A. □

DEFINITION 11.3.12. Let B/A be an extension of commutative rings. The *integral closure* of A in B is the set of elements of B that are integral over A.

PROPOSITION 11.3.13. *Let B/A be an extension of commutative rings. Then the integral closure of A in B is a subring of B.*

PROOF. If α and β are elements of B that are integral over A, then $A[\alpha,\beta]$ is integral over A by Proposition 11.3.10. Therefore, every element of $A[\alpha,\beta]$, including $-\alpha$, $\alpha+\beta$, and $\alpha\cdot\beta$, is integral over A as well. That is, the integral closure of A in B is closed under addition, additive inverses, and multiplication, and it contains 1, so it is a ring. □

EXAMPLE 11.3.14. The integral closure of $\mathbb{Z}$ in $\mathbb{Z}[x]$ is $\mathbb{Z}$, since if $f \in \mathbb{Z}[x]$ is of degree at least 1 and $g \in \mathbb{Z}[x]$ is nonconstant, then $g(f(x))$ has degree $\deg g \cdot \deg f$ in x, hence cannot be 0.

DEFINITION 11.3.15.

a. The *ring of algebraic integers* is the integral closure $\overline{\mathbb{Z}}$ of $\mathbb{Z}$ inside $\mathbb{C}$.

b. An *algebraic integer* is an element of $\overline{\mathbb{Z}}$.

DEFINITION 11.3.16. Let B/A be an extension of commutative rings. We say that A is *integrally closed* in B if A is its own integral closure in B.

DEFINITION 11.3.17. We say that an integral domain A is *integrally closed*, or , if it is integrally closed in its quotient field.

EXAMPLE 11.3.18. Every field is integrally closed.

PROPOSITION 11.3.19. *Let A be an integrally closed domain, let K be the quotient field, and let L be a field extension of K. If $\beta \in L$ is integral over A with minimal polynomial $f \in K[x]$, then $f \in A[x]$.*

PROOF. Since $\beta \in L$ is integral, it is the root of some monic polynomial $g \in A[x]$ such that f divides g in $K[x]$. As g is monic, every root of g in an algebraic closure $\overline{K}$ containing K is integral over K. As every root of f is a root of g, the same is true of the roots of f. Write $f = \prod_{i=1}^{n}(x-\beta_i)$ for $\beta_i \in \overline{K}$ integral over A. As the integral closure of A in $\overline{K}$ is a ring, it follows that every coefficient of f is integral over A, being sums of products of the elements β_i. Since $f \in K[x]$ and A is integrally closed, we then have $f \in A[x]$. □

In particular, we have the following result on the norm and trace on quotient fields on integral extensions of domains.

COROLLARY 11.3.20. *Let B/A be an integral extension of domains, and suppose that A is integrally closed in its quotient field K. Let L denote the quotient field of B, and suppose that L/K is finite. Then $N_{L/K}(\beta)$ and $\mathrm{Tr}_{L/K}(\beta)$ are elements of A for every $\beta \in B$.*

The following holds in the case of UFDs.

PROPOSITION 11.3.21. *Let A be a UFD, let K be the quotient field of A, and let L be a field extension of K. Suppose that $\beta \in L$ is algebraic over K with minimal polynomial $f \in K[x]$. If β is integral over A, then $f \in A[x]$.*

PROOF. Let $\beta \in L$ be integral over A, let $g \in A[x]$ be a monic polynomial of which it is a root, and let $f \in K[x]$ be the minimal polynomial of β. Since f divides g in $K[x]$ and A is a UFD with quotient field K, there exists $d \in K$ such that $df \in A[x]$ and df divides g in $A[x]$. Since f is monic, d must be an element of A (and in fact may be taken to be a least common denominator of the coefficients of f). The coefficient of the leading term of any multiple of df will be divisible by d, so this forces d to be a unit, in which case $f \in A[x]$. □

COROLLARY 11.3.22. *Every unique factorization domain is integrally closed.*

PROOF. The minimal polynomial of an element a of the quotient field K of a UFD A is $x-a$. If $a \notin A$, it follows from Proposition 11.3.21 that a is not integral over A. □

EXAMPLES 11.3.23. The ring $\mathbb{Z}$ is integrally closed.

EXAMPLE 11.3.24. The ring $\mathbb{Z}[\sqrt{17}]$ is not integrally closed, since $\alpha = \frac{1+\sqrt{17}}{2}$ is a root of the monic polynomial x^2-x-4. In particular, $\mathbb{Z}[\sqrt{17}]$ is not a UFD.

PROPOSITION 11.3.25. *Let B/A be an extension of commutative rings, and suppose that B is an integrally closed domain. Then the integral closure of A in B is integrally closed.*

PROOF. Let $\overline{A}$ denote the integral closure of A in B, and let Q denote the quotient field of $\overline{A}$. Let $\alpha \in Q$, and suppose that α is integral over $\overline{A}$. Then $\overline{A}[\alpha]$ is integral over $\overline{A}$, so $\overline{A}[\alpha]$ is integral over A, and therefore α is integral over A. That is, α is an element of $\overline{A}$, as desired. □

EXAMPLE 11.3.26. The ring $\overline{\mathbb{Z}}$ of algebraic integers is integrally closed.

PROPOSITION 11.3.27. *Let A be an integral domain with quotient field K, and let L be an algebraic extension of K. Then the integral closure B of A in L has quotient field equal to L inside L. In fact, every element of L may be written as $\frac{b}{d}$ for some $d \in A$ and $b \in B$.*

PROOF. Any $\beta \in L$ is the root of a nonconstant polynomial $f = \sum_{i=0}^{n} a_i x^i \in K[x]$, with $a_n = 1$. Let $d \in A$ be such that $df \in A[x]$. Then

$$d^n f(d^{-1}x) = \sum_{i=0}^{n} a_i d^{n-i} x^i \in A[x]$$

is both monic and has $d\beta$ as a root. In other words, $d\beta$ is contained in B, as desired. □

EXAMPLE 11.3.28. The quotient field of $\overline{\mathbb{Z}}$ is $\overline{\mathbb{Q}}$.

DEFINITION 11.3.29. A *number field* (or *algebraic number field*) is a finite field extension of $\mathbb{Q}$.

We have the following names for extensions of $\mathbb{Q}$ of various degrees.

DEFINITION 11.3.30. A *quadratic* (resp., *cubic, quartic, quintic*, ...) *field* is a degree 2 (resp., 3, 4, 5, ...) extension of $\mathbb{Q}$.

DEFINITION 11.3.31. The *ring of integers* (or *integer ring*) of a number field K is the integral closure of $\mathbb{Z}$ in K.

In other words, the ring of integers of a number field is the subring of algebraic integers it contains. The prototypical examples of rings of integers arise in the setting of quadratic fields.

THEOREM 11.3.32. *Let $d \neq 1$ be a square-free integer. The ring $\mathscr{O}$ of algebraic integers in $\mathbb{Q}(\sqrt{d})$ is*

$$\mathscr{O} = \begin{cases} \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right] & \textit{if } d \equiv 1 \bmod 4, \\ \mathbb{Z}[\sqrt{d}] & \textit{if } d \equiv 2,3 \bmod 4. \end{cases}$$

PROOF. Suppose that $\alpha = a + b\sqrt{d}$ is integral for $a, b \in \mathbb{Q}$. If $b = 0$, then we must have $a \in \mathbb{Z}$. If $b \neq 0$, then the minimal polynomial of α is $f = x^2 - 2ax + a^{?}?2 - b^2d$. Since α is integral, then we must have $f \in \mathbb{Z}[x]$, so $2a \in \mathbb{Z}$. If $a \in \mathbb{Z}$, then since $a^2 - b^2d \in \mathbb{Z}$ and d is square-free, we have $b \in \mathbb{Z}$ as well. If $a \notin \mathbb{Z}$, then $2a = a'$ and $2b = b'$ for some odd $a', b' \in \mathbb{Z}$, and $(a')^2 \equiv (b')^2 d \bmod 4$. As $(\mathbb{Z}/4\mathbb{Z})^2 = \{0, 1\}$, this is impossible if $d \not\equiv 1 \bmod 4$. If $d \equiv 1 \bmod 4$, then clearly we can take $a' = b' = 1$. □

DEFINITION 11.3.33. Let B/A be an integral extension of domains such that A is integrally closed, and suppose that B is free of rank n as an A-module. Let $(\beta_1, \ldots, \beta_n)$ be an ordered basis of B as a free A-module. The *discriminant* B over A relative to the basis $(\beta_1, \ldots, \beta_n)$ is $\mathrm{D}(\beta_1, \ldots, \beta_n)$.

LEMMA 11.3.34. *Let A be an integrally closed domain with quotient field K. Let L be a finite separable extension of K, and let B denote the integral closure of A in L. Let* $(\alpha_1, \ldots, \alpha_n)$

be any ordered basis of L as a K-vector space that is contained in B. Let $\beta \in L$ be such that $\mathrm{Tr}_{L/K}(\alpha\beta) \in A$ for all $\alpha \in B$. Then

$$\mathrm{D}(\alpha_1,\dots,\alpha_n)\beta \in \sum_{i=1}^{n} A\alpha_i.$$

PROOF. Since $\beta \in L$, we may write

$$\beta = \sum_{i=1}^{n} a_i\alpha_i$$

for some $a_i \in K$ for $1 \le i \le n$. For any i, we have that

$$\mathrm{Tr}_{L/K}(\alpha_i\beta) = \sum_{j=1}^{n} a_j \,\mathrm{Tr}_{L/K}(\alpha_i\alpha_j). \tag{11.3.1}$$

The right-hand side of (11.3.1) is the ith term of the product of the matrix $Q = (\mathrm{Tr}_{L/K}(\alpha_i\alpha_j))$ times the column vector with ith entry a_i. Since the determinant of Q is $d = \mathrm{D}(\alpha_1,\dots,\alpha_n)$, letting $Q^* \in M_n(A)$ denote the adjoint matrix to Q, we have $Q^*Q = dI_n$. Thus, we have $da_i \in A$ for each i. In other words, $d\beta$ lies in the A-module generated by the α_i, so we are done. □

PROPOSITION 11.3.35. *Let A be an integrally closed domain with quotient field K. Let L be a finite separable extension of K, and let B denote the integral closure of A in L. There exists an ordered basis $(\alpha_1,\dots,\alpha_n)$ of L as a K-vector space contained in B. Moreover, for any such basis, we have*

$$\sum_{i=1}^{n} A\alpha_i \subseteq B \subseteq \sum_{i=1}^{n} Ad^{-1}\alpha_i,$$

where $d = \mathrm{D}(\alpha_1,\dots,\alpha_n)$.

PROOF. First, take any ordered basis $(\beta_1,\dots,\beta_n)$ of L/K. By Proposition 11.3.27, there exists $a \in A - \{0\}$ such that $\alpha_i = a\beta_i \in B$ for each $1 \le i \le n$. Clearly, $(\alpha_1,\dots,\alpha_n)$ is a basis of L/K, so in particular, the A-module generated by the α_i is free and contained in B. The other containment is simply a corollary of Lemma 11.3.34 and the fact that $\mathrm{Tr}_{L/K}(B) \subseteq A$. □

The following notion of rank is most interesting for finitely generated modules, though we shall have occasion to use it without this assumption.

DEFINITION 11.3.36. The *rank* of a module M over a domain A is

$$\mathrm{rank}_A(M) = \dim_K(K \otimes_A M).$$

COROLLARY 11.3.37. *Let A be an integrally closed noetherian domain with quotient field K. Let L be a finite separable extension of K, and let B denote the integral closure of A in L. Then B is a finitely generated, torsion-free A-module of rank $[L : K]$.*

PROOF. By Proposition 11.3.35, we have free A-modules M and M' of rank $n = [L : K]$ such that $M \subseteq A \subseteq M'$. Since M' has no A-torsion, neither does B. We have

$$K \otimes_A M \subseteq K \otimes_A B \subseteq K \otimes_A M'.$$

As M and M' are both isomorphic to A^n, their tensor products over A with K are n-dimensional K-vector spaces, which forces $K \otimes_A B$ to have K-dimension n as well. Moreover, B is finitely generated being a submodule of a finitely generated module over A, as A is noetherian. □

PROPOSITION 11.3.38. *Let A be an integrally closed noetherian domain with quotient field K. Let L be a finite separable extension of K, and let B denote the integral closure of A in L. Then any finitely generated, nonzero B-submodule of L is a torsion-free A-module of rank $[L:K]$.*

PROOF. Let M be a finitely generated, nonzero B-submodule of L. If $\beta \in L^\times$, then the multiplication-by-β map $B \to B\beta$ is an isomorphism of B-modules, so $B\beta$ has rank $[L:K]$ as an A-module. In particular, $\mathrm{rank}_A(M) \ge \mathrm{rank}_A(B)$, taking $\beta \in M$. Since M is B-finitely generated and contained in the quotient field of B, there exists $\alpha \in B$ such that $\alpha M \subseteq B$. Since multiplication by α is an isomorphism, $\mathrm{rank}_A(M) \le \mathrm{rank}_A(B)$. The result now follows from Corollary 11.3.37. □

COROLLARY 11.3.39. *Let A be a PID with quotient field K, let L be a finite separable extension of K, and let B denote the integral closure of K in L. Then any finitely generated B-submodule of L is a free A-module of rank $[L:K]$.*

PROOF. By the structure theorem for modules over a PID, any torsion-free rank n module over A is isomorphic to A^n. The result is then immediate from Proposition 11.3.38. □

We have the following application to number fields.

LEMMA 11.3.40. *Let K be a number field. Then the discriminant of $\mathscr{O}_K$ over $\mathbb{Z}$ is independent of the choice of ordered basis of $\mathscr{O}_K$ as a free $\mathbb{Z}$-module.*

PROOF. By Corollary 11.3.39, the ring $\mathscr{O}_K$ is free of rank $n = [K:\mathbb{Q}]$ over $\mathbb{Z}$. If $\beta_1, \dots, \beta_n$ and $\alpha_1, \dots, \alpha_n$ are bases of $\mathscr{O}_K$ as a free $\mathbb{Z}$-module, then there exists a $\mathbb{Q}$-linear homomorphism $T\colon K \to K$ such that $T(\alpha_i) = \beta_i$ for all i. Then

$$\mathrm{D}(\beta_1, \dots, \beta_n) = \det(T)^2 \mathrm{D}(\alpha_1, \dots, \alpha_n),$$

and $\det(T)$ is a unit in $\mathbb{Z}$, so in $\{\pm 1\}$, which is to say that $\det(T)^2 = 1$. □

DEFINITION 11.3.41. If K is a number field, the *discriminant* $\mathrm{disc}(K)$ of K is the discriminant of $\mathscr{O}_K$ over $\mathbb{Z}$ relative to any basis of $\mathscr{O}_K$ as a free $\mathbb{Z}$-module.

Noting Theorem 11.3.32, the case of quadratic fields is immediately calculated.

PROPOSITION 11.3.42. *Let $K = \mathbb{Q}(\sqrt{d})$, where $d \neq 1$ is a square-free integer. Then*

$$\mathrm{disc}(K) = \begin{cases} d & d \equiv 1 \bmod 4, \\ 4d & d \equiv 2, 3 \bmod 4. \end{cases}$$

The following theorem will be useful to us later.

THEOREM 11.3.43 (Noether's normalization lemma). *Let F be a field, and let A be a finitely generated commutative F-algebra with generators $z_1, \dots, z_r \in A$. Then there exists $s \le r$ and F-algebraically independent elements $t_1, \dots, t_s \in A$ such that A is integral over $F[t_1, \dots, t_s]$.*

PROOF. The result is obvious for $r = 0$, so suppose $r \geq 1$. If the elements $z_1, \dots, z_r$ are algebraically independent over F, then we may take $s = r$ and $t_i = r_i$ for all i, so suppose not. In this case, there exists a nonzero polynomial $f \in F[x_1, \dots, x_r]$ such that $f(z_1, \dots, z_r) = 0$. Let d be the maximum of the degrees of f viewed as a polynomial in each of the x_i. Since f is nonconstant, without loss of generality we may take it to be nonconstant as a polynomial in x_1 with coefficients in $F[x_2, \dots, x_r]$.

Consider the polynomial

$$g(x_1, \dots, x_r) = f\left(x_1, x_2 + x_r^{d+1}, \dots, x_r + x_r^{(d+1)^{r-1}}\right).$$

Each monomial $x_1^{k_1} \cdots x_r^{k_r}$ in f has $k_i \leq d$ for all i and gives rise to a sum of monomials in g, exactly one of which has the form a constant in F times x_1 to the power $\sum_{i=1}^r k_i(1+d)^{i-1}$. Each of these powers for the different monomials in f is distinct, so the highest degree term in g viewed as a polynomial in x_1 has a nonzero coefficient c that lies in F. That is, $c^{-1}g$ is monic as a polynomial in x_1 with coefficients in $F[x_2, \dots, x_r]$.

Set

$$w_i = z_i - z_1^{(d+1)^{i-1}}$$

for $2 \leq i \leq r$, and note that $g(z_1, w_2, \dots, w_r) = f(z_1, \dots, z_r) = 0$. It follows that z_r is integral over $B = F[w_2, \dots, w_r]$. By induction, there exist $s \leq r$ and elements $t_1, \dots, t_s \in B$ that are algebraically independent over F and for which B is integral over $F[t_1, \dots, t_s]$. Then $A = B[z_1]$ is integral over $F[t_1, \dots, t_s]$ by the transitivity of integral extensions, proving the theorem. □

COROLLARY 11.3.44. *Let K be an extension of a field F that is finitely generated as an F-algebra. Then K is a finite extension of F.*

PROOF. By Noether's normalization lemma, L is an integral extension of $F[t_1, \dots, t_s]$ for some algebraically independent elements $t_1, \dots, t_s \in L$. However, L is a field so contains the quotient field $K(t_1, \dots, t_s)$. Since no t_i^{-1} is integral over $F[t_1, \dots, t_s]$, we must have $s = 0$. Thus L is integral over F, which is to say it is an algebraic extension of F, but then it is clearly finite being that it is generated by finitely many elements. □

11.4. Radicals of ideals

Let R be a commutative ring.

DEFINITION 11.4.1. The *radical* $\sqrt{I}$ of an ideal I of R is the set

$$\sqrt{I} = \{a \in R \mid a^k \in I \text{ for some } k \geq 1\}.$$

LEMMA 11.4.2. *For any ideal I of R, the radical $\sqrt{I}$ of I is an ideal of R.*

PROOF. If $r \in R$ and $a \in \sqrt{I}$, then there exists $k \geq 1$ with $a^k \in I$, and then $(ra)^k = r^k a^k \in I$, so $ra \in \sqrt{I}$ as well. If we also have $b \in \sqrt{l}$ with $b^l \in I$, then

$$(a+b)^{k+l} = \sum_{i=0}^{k+l} \binom{k+l}{i} a^i b^{k+l-i} \in I$$

since either $i \geq k$ or $k + l - i \geq l$ if $i \in \mathbb{Z}$. Thus $a + b \in \sqrt{I}$. □

The nilradical of R is the radical of the ideal (0) of R.

DEFINITION 11.4.3. An element $a \in R$ is *nilpotent* if there exists $n \geq 1$ such that $a^n = 0$.

DEFINITION 11.4.4. The *nilradical* of R is the ideal of nilpotent elements in R.

EXAMPLE 11.4.5. The nilradical of $F[x]/(x^n)$ for $n \geq 1$ is generated by x.

In fact, the following is easily verified.

LEMMA 11.4.6. *If $\pi \colon R \to R/I$ is the projection of R onto its quotient by an ideal I, then $\pi(\sqrt{I})$ is the nilradical of R/I.*

We leave the following as an exercise that uses Zorn's lemma.

PROPOSITION 11.4.7. *Let I be a proper ideal of R. Then $\sqrt{I}$ is the intersection of all prime ideals of R containing I.*

In particular, the radical of a prime ideal is itself.

DEFINITION 11.4.8. An ideal is *radical*, or a *semiprime ideal*, if it is its own radical.

EXAMPLES 11.4.9.

a. Prime ideals are radical.

b. Let F be a field, $f_1, \ldots, f_r \in F[x]$ be irreducible, and $k_1, \ldots, k_r$ be positive integers. Then

$$\sqrt{(f_1^{k_1} \cdots f_r^{k_r})} = (f_1 \cdots f_r).$$

Thus, the nonzero radical ideals of $F[x]$ are exactly the ideals generated by products of distinct irreducible elements.

c. Radicals of ideals are radical.

PROPOSITION 11.4.10. *Let R be noetherian and I be an ideal of R. Then there exists $N \geq 1$ such that $(\sqrt{I})^n \subseteq I$ for all $n \geq N$.*

PROOF. Let $a_1, \ldots, a_m \in R$ be such that $\sqrt{I} = (a_1, \ldots, a_m)$. For $1 \leq i \leq m$, let $k_i \geq 1$ be such that $a_i^{k_i} \in I$, and let $k = \max\{k_i \mid 1 \leq i \leq m\}$. For any $x = \sum_{i=1}^m r_i a_i \in \sqrt{I}$, we have

$$x^{km} \in (\{a_1^{i_1} \cdots a_m^{i_m} \mid i_j \geq 0 \text{ for all } j \text{ with } i_1 + \cdots + i_m = km\}) \subseteq (a_1^k, \ldots, a_r^k) \subseteq I.$$

□

DEFINITION 11.4.11. An ideal I of R is *nilpotent* if there exists $n \geq 1$ such that $I^n = 0$.

COROLLARY 11.4.12. *The nilradical of a noetherian commutative ring is nilpotent.*

It is easy to see why this can fail in a noncommutative ring.

EXAMPLE 11.4.13. Consider the polynomial ring $R = F[x_1, x_2, \ldots]$ in countably many variables over a field F and its ideal $I = (x_k^k \mid k \geq 1)$. Its radical is $\sqrt{I} = (x_k \mid k \geq 1)$ but no power of $\sqrt{I}$ is contained in I.

The taking of radicals behaves well with respect to localization.

LEMMA 11.4.14. *Let S be a multiplicatively closed subset of R, and let I be an ideal of R. Then $S^{-1}\sqrt{I} = \sqrt{S^{-1}I}$.*

11.5. Going up and going down

We use B/A to denote an extension of commutative rings.

DEFINITION 11.5.1. Let B/A be an extension of commutative rings. We say that an ideal $\mathfrak{b}$ of B *lies over* an ideal $\mathfrak{a}$ of A if $\mathfrak{b} \cap A = \mathfrak{a}$.

We begin by noting the following simple lemma.

LEMMA 11.5.2. *Let A be an integral domain, and let B be a commutative ring extension of A that is integral over A. If $\mathfrak{b}$ is an ideal of B that contains a nonzero element which is not a zero divisor, then $\mathfrak{b}$ lies over a nonzero ideal of A.*

PROOF. That $\mathfrak{b} \cap A$ is an ideal is clear, so it suffices to show that $\mathfrak{b} \cap A$ is nonzero. Let $\beta \in \mathfrak{b}$ be nonzero and not a zero divisor. Then β is a root of some monic polynomial $g \in A[x]$. Write $g = x^n f$ for some nonzero $f \in A[x]$ with nonzero constant term. Since $\beta \in \mathfrak{b}$, we have $f(\beta) - f(0) \in \mathfrak{b}$, and as $f(\beta) = 0$ given that β is not a zero divisor, we have $f(0) \in \mathfrak{b}$. But $f(0) \neq 0$, so $\mathfrak{b}$ has a nonzero element. □

The following are also easily verified.

LEMMA 11.5.3. *If B/A is integral and $\mathfrak{b}$ is an ideal of B that lies over $\mathfrak{a}$, then $B/\mathfrak{b}$ is integral over $A/\mathfrak{a}$.*

LEMMA 11.5.4. *Let S be a multiplicatively closed subset of A. If B/A is integral, then so is $S^{-1}B/S^{-1}A$.*

PROPOSITION 11.5.5. *Let B/A be an integral extension. If $\mathfrak{p}$ is a prime ideal of A, then there exists a prime ideal $\mathfrak{q}$ of B lying over $\mathfrak{p}$.*

PROOF. Let $B_{\mathfrak{p}} = S_{\mathfrak{p}}^{-1}B$ be integral over $A_{\mathfrak{p}}$. Let $\mathfrak{M}$ be a maximal ideal of $B_{\mathfrak{p}}$. Then $\mathfrak{m} = \mathfrak{M} \cap A_{\mathfrak{p}}$ is maximal, since $A_{\mathfrak{p}}/\mathfrak{m}$ injects into the field $B_{\mathfrak{p}}/\mathfrak{M}$. Since $A_{\mathfrak{p}}$ is local, we have $\mathfrak{m} = \mathfrak{p}A_{\mathfrak{p}}$. Let $\iota \colon B \to B_{\mathfrak{p}}$ be the localization map so that $\mathfrak{q} = \iota^{-1}(\mathfrak{M})$ is prime, and $\mathfrak{q} \cap A = \iota^{-1}(\mathfrak{p}A_{\mathfrak{p}}) = \mathfrak{p}$. □

THEOREM 11.5.6 (Going up). *Let B/A be an integral extension. Suppose that $\mathfrak{p}_1 \subseteq \mathfrak{p}_2$ are prime ideals of A and $\mathfrak{q}_1$ is a prime ideal of B lying over $\mathfrak{p}_1$. Then there exists a prime ideal $\mathfrak{q}_2$ of B containing $\mathfrak{q}_1$ and lying over $\mathfrak{p}_2$.*

PROOF. Let $\bar{A} = A/\mathfrak{p}_1$ and $\bar{B} = B/\mathfrak{p}_2$, and let $\pi \colon B \to \bar{B}$ be the quotient map. Let $\bar{\mathfrak{p}}_2$ be the image of $\mathfrak{p}_2$ in $\bar{A}$. By Proposition 11.5.5, there exists a prime ideal $\bar{\mathfrak{q}}_2$ of $\bar{B}$ lying over $\bar{\mathfrak{p}}_2$. Then $\mathfrak{q}_2 = \pi^{-1}(\bar{\mathfrak{q}}_2)$ contains $\mathfrak{q}_1$ and satisfies

$$\mathfrak{q}_2 \cap A = \pi^{-1}(\bar{\mathfrak{q}}_2 \cap \bar{A}) = \pi^{-1}(\bar{\mathfrak{p}}_2) = \mathfrak{p}_2,$$

since $\mathfrak{p}_2$ contains $\mathfrak{p}_1$. □

The proof of the following two propositions are left as exercises.

PROPOSITION 11.5.7. *Let B/A be an integral extension of domains, and let B' be the integral closure of A in B. Let $\mathfrak{a}$ be an ideal of A. The following conditions on $\beta \in B$ are equivalent: An element $\beta \in B$ is a root of $f \in A[x]$ with $f = x^n + \sum_{i=0}^{n-1} a_i x^i$ with $a_i \in \mathfrak{a}$ for all $0 \le i \le n-1$ if and only if $\beta \in \sqrt{\mathfrak{a}B'}$.*

PROPOSITION 11.5.8. *Let B/A be an integral extension of domains such that A is integrally closed, and suppose that $\beta \in B$ is the root of a monic polynomial in $A[x]$ with non-leading coefficients in an ideal $\mathfrak{a}$ of A. Then the minimal polynomial of β is also such a polynomial.*

LEMMA 11.5.9. *Let $\mathfrak{p}$ be a prime ideal of A. There exists a prime ideal $\mathfrak{q}$ of B lying over $\mathfrak{p}$ if and only if $\mathfrak{p}B \cap A = \mathfrak{p}$.*

PROOF. If $\mathfrak{q} \cap A = \mathfrak{p}$, then $\mathfrak{q}$ contains $\mathfrak{p}B$, and then

$$\mathfrak{p} = \mathfrak{q} \cap A \supseteq \mathfrak{p}B \cap A \supseteq \mathfrak{p},$$

so $\mathfrak{p}B \cap A = \mathfrak{p}$.

Conversely, if $\mathfrak{p}B \cap A = \mathfrak{p}$, then $\mathfrak{p}B$ is disjoint from $S_{\mathfrak{p}}$, so there exists a maximal ideal $\mathfrak{M}$ of $B_{\mathfrak{p}}$ with $\mathfrak{p}B$ contained in $\mathfrak{M}$. Let $\mathfrak{q}$ be the inverse image of $\mathfrak{M}$ in A. Then $\mathfrak{m} = \mathfrak{q} \cap A$ is a prime ideal containing $\mathfrak{p}$ with $\mathfrak{m} \cap S_{\mathfrak{p}} = \varnothing$, which forces $\mathfrak{m} = \mathfrak{p}$. □

THEOREM 11.5.10 (Going down). *Let B/A be an integral extension of integral domains with A integrally closed. Suppose that $\mathfrak{p}_2 \subseteq \mathfrak{p}_1$ are prime ideals of A and $\mathfrak{q}_1$ is a prime ideal of B lying over $\mathfrak{p}_1$. Then there exists a prime ideal $\mathfrak{q}_2$ of B contained in $\mathfrak{q}_1$ and lying over $\mathfrak{p}_2$.*

PROOF. Note that the maps $B \to B_{\mathfrak{q}_1}$ and $A \to A_{\mathfrak{p}_1}$ are injective as B is a domain. By Lemma 11.5.9, it is enough to show that $\mathfrak{p}_2 B_{\mathfrak{q}_1} \cap A = \mathfrak{p}_2$. That is, in this case there exists a prime ideal $\mathfrak{Q}_2$ of $B_{\mathfrak{q}_1}$ lying over $\mathfrak{p}_2$, and then we can take $\mathfrak{q}_2 = \mathfrak{Q}_2 \cap A$.

If $\beta = \frac{b}{s} \in \mathfrak{p}_2 B_{\mathfrak{q}_1}$ with $b \in \mathfrak{p}_2 B$ and $s \in B - \mathfrak{q}_1$, then by Proposition 11.5.8, the minimal polynomial $f = x^n + \sum_{i=0}^{n-1} a_i x^i$ of b has non-leading coefficients in $\mathfrak{p}_2$. If β is also in A, then $s = \beta^{-1} b$ has minimal polynomial $\beta^{-n} f(\beta x) \in Q(A)[x]$. Since s is integral over A, this polynomial lies in $A[x]$, and therefore $\beta^{i-n} a_i \in A$ for all i. If $\beta \notin \mathfrak{p}_2$, then for all i we have $\beta^{i-n} a_i \in \mathfrak{p}_2$ since $a_i \in \mathfrak{p}_2$. But then $s^n \in \mathfrak{p}_2 B$, so $s \in \mathfrak{q}_1$, a contradiction. Thus, $\beta \in \mathfrak{p}_2$, as required. □

11.6. Primary decomposition

DEFINITION 11.6.1. A proper ideal $\mathfrak{q}$ of R is *primary* if for any $a, b \in R$ with $ab \in \mathfrak{q}$, one has either $a \in \mathfrak{q}$ or $b^n \in \mathfrak{q}$ for some $n \ge 1$.

That is, an ideal $\mathfrak{q}$ is primary if whenever $ab \in \mathfrak{q}$, either $a \in \mathfrak{q}$ or $b \in \sqrt{\mathfrak{q}}$. Of course, prime ideals are primary. The following is just a rephrasing of the definition of primary.

LEMMA 11.6.2. *A proper ideal $\mathfrak{q}$ of R is primary if and only if every zero divisor in $R/\mathfrak{q}$ is nilpotent.*

The following is a key property of primary ideals.

PROPOSITION 11.6.3. *The radical of any primary ideal is a prime ideal.*

PROOF. Let $\mathfrak{q}$ be a primary ideal of R. If $a,b \in R$ with $ab \in \sqrt{\mathfrak{q}}$, then $a^k b^k \in \mathfrak{q}$ for some $k \geq 1$, and therefore either $a^k \in \mathfrak{q}$ or there exists $n \geq 1$ such that $b^{kn} \in \mathfrak{q}$. In the first, case $a \in \sqrt{\mathfrak{q}}$, and in the second, $b \in \sqrt{\mathfrak{q}}$, so $\sqrt{\mathfrak{q}}$ is prime. □

In particular, the radical of a primary ideal $\mathfrak{q}$ is the smallest prime ideal of R containing $\mathfrak{q}$, given that it is also the intersection of all prime ideals containing $\mathfrak{q}$.

DEFINITION 11.6.4. The radical $\mathfrak{p}$ of a primary ideal $\mathfrak{q}$ of R is called the *associated prime* to $\mathfrak{q}$, and we say that $\mathfrak{q}$ is $\mathfrak{p}$-*primary*.

EXAMPLES 11.6.5. Let F be a field.

a. The ideal (x^2,y) of $F[x,y]$ is primary since $F[x,y]/(x^2,y) \cong F[x]/(x^2)$, and every zero divisor in the latter ring is nilpotent. Its associated prime is (x,y).

b. Consider $R = F[x,y,z]/(xy - z^2)$ and its ideal $\mathfrak{p} = (x,z)$, which is prime since $R/\mathfrak{p} \cong F[y]$. We have $xy \in \mathfrak{p}^2$, but $x \notin \mathfrak{p}^2$ and $y \notin \sqrt{\mathfrak{p}^2} = \mathfrak{p}$. Thus $\mathfrak{p}^2$ is not primary, even though $\mathfrak{p}$ is prime.

LEMMA 11.6.6. *If I is an ideal of R such that $\sqrt{I}$ is maximal, then I is primary. In particular, any power of a maximal ideal $\mathfrak{m}$ is primary with associated prime $\mathfrak{m}$.*

PROOF. Suppose that $\mathfrak{m} = \sqrt{I}$ is maximal. The image of $\mathfrak{m}$ in R/I is the nilradical of R/I, which means that the nilradical is the only prime ideal of R/I. That is, $\mathfrak{m}$ is local, and every element of R/I that is not nilpotent is a unit. In particular, every zero divisor of R/I is nilpotent. Thus, I is $\mathfrak{m}$-primary. □

The following is easily checked.

LEMMA 11.6.7. *A finite intersection of primary ideals with the same associated prime is primary.*

DEFINITION 11.6.8. Let I be an ideal of R.

a. A *primary decomposition* of I is a finite collection $\{\mathfrak{q}_1,\dots,\mathfrak{q}_k\}$ of primary ideals of R such that $I = \bigcap_{i=1}^k \mathfrak{q}_i$.

b. We say that I is *decomposable* if it has a primary decomposition.

c. A primary decomposition $\{\mathfrak{q}_1,\dots,\mathfrak{q}_k\}$ of I is *minimal* if the radicals $\sqrt{\mathfrak{q}_i}$ are all distinct and no proper subset of the primary decomposition is also a primary decomposition of I.

Every ideal with a primary decomposition has a minimal such decomposition.

LEMMA 11.6.9. *Let I be a decomposable ideal of R. Then I has a minimal primary decomposition.*

PROOF. From this decomposition, we may first remove one at a time any primes that contain the intersection of the others. By Lemma 11.6.7, we may then replace the subcollection of those ideals in the decomposition with the same associated prime by the single primary ideal that is its intersection. The resulting collection is minimal. □

DEFINITION 11.6.10. A proper ideal I of R is *irreducible* if for any ideals $\mathfrak{a}$ and $\mathfrak{b}$ of R with $I = \mathfrak{a} \cap \mathfrak{b}$, either $I = \mathfrak{a}$ or $I = \mathfrak{b}$.

PROPOSITION 11.6.11. *Let R be noetherian. Then every irreducible ideal of R is primary.*

PROOF. Let I be an irreducible ideal of R, and let $a,b \in R$ with $ab \in I$ but $b \notin I$. For each $n \geq 1$, let $J_n = \{r \in R \mid a^n r \in I\}$, and note that J_n is an ideal of R. Then J_n form an ascending chain of ideals containing I, and since R is noetherian, this chain is eventually constant, say $J_n = J_{n+1}$ for all $n \geq N$ with $N \geq 1$. Consider the ideals $\mathfrak{a} = (a^N) + I$ and $\mathfrak{b} = (b) + I$ containing I. We claim that $\mathfrak{a} \cap \mathfrak{b} = I$. Let $c \in \mathfrak{a} \cap \mathfrak{b}$. Then $c = a^N r + q$ for some $r \in R$ and $q \in I$. Since $c \in (b) + I$, we have $ac \in (ab) + I = I$. In other words, $a^{N+1}r + qa \in I$, so $a^{N+1}r \in I$, so $r \in J_{N+1} = J_N$. Therefore $a^N r \in I$ as well, so $c \in I$, and the claim holds. Since I is irreducible and $b \notin I$, we must have $I = \mathfrak{a}$, which means that $a^N \in I$. Therefore, I is primary. □

PROPOSITION 11.6.12. *Let R be noetherian. The every proper ideal of R is a finite intersection of irreducible ideals.*

PROOF. Let X be the set of proper ideals of R that cannot be written as a finite intersection of irreducible ideals of R. Since R is noetherian, either X is empty or X has a maximal element $\mathfrak{m}$. Since $\mathfrak{m} \in X$, it is not irreducible, so there exist ideals $\mathfrak{a}$ and $\mathfrak{b}$ properly containing $\mathfrak{m}$ with $\mathfrak{m} = \mathfrak{a} \cap \mathfrak{b}$. Since $\mathfrak{m}$ is maximal in X, both of $\mathfrak{a}$ and $\mathfrak{b}$ can be written as a finite intersection of irreducible ideals, so $\mathfrak{m}$ may be as well, which contradicts the existence of $\mathfrak{m}$. Therefore X is empty, as desired. □

Combining Propositions 11.6.11 and 11.6.12, we have the following.

THEOREM 11.6.13 (Primary decomposition theorem). *Every proper ideal of a noetherian commutative ring R is decomposable.*

We now consider uniqueness of primary decompositions, given the existence of one. We begin with the following simple lemma.

LEMMA 11.6.14.

a. Let $\mathfrak{p}_1, \ldots, \mathfrak{p}_k$ be prime ideals of R. If an ideal I is contained in $\bigcup_{i=1}^k \mathfrak{p}_i$, then I is contained in some $\mathfrak{p}_i$.

b. Let $\mathfrak{a}_1, \ldots, \mathfrak{a}_k$ be ideals of R. If a prime ideal $\mathfrak{p}$ contains (resp., equals) $\bigcap_{i=1}^k \mathfrak{a}_i$, then $\mathfrak{p}$ contains (resp., equals) some $\mathfrak{a}_i$.

PROOF. We prove part a by induction on k, it being clearly true for $k = 1$. Suppose that I is not contained in any $\mathfrak{p}_i$ but I is contained in the union of the $\mathfrak{p}_i$. By induction, for each i, we can find $a_i \in I$ such that $a_i \notin \mathfrak{p}_j$ for all $j \neq i$. By assumption, we then have $a_i \in \mathfrak{p}_i$ for each i. The element

$$b = \sum_{i=1}^{k} \prod_{\substack{j=1 \\ j \neq i}}^{k} a_i$$

of I has image in $R/\mathfrak{p}_i$ equal to the image of its ith term, which is nonzero by the primality of i. That is, $b \notin \bigcup_{i=1}^k \mathfrak{p}_i$, which is a contradiction.

As for part b, let $\mathfrak{a} = \bigcap_{i=1}^k \mathfrak{a}_i$. Suppose that $\mathfrak{p}$ does not contain any $\mathfrak{a}_i$, and choose $a_i \in \mathfrak{a}_i$ with $a_i \notin \mathfrak{p}$ for each i. Then $a = \prod_{i=1}^k a_i \in \mathfrak{a}$, but $a \notin \mathfrak{p}$ by primality of $\mathfrak{p}$. Therefore, $\mathfrak{p}$ does not contain

$\mathfrak{a}$. If on the other hand $\mathfrak{p} = \mathfrak{a}$, then $\mathfrak{p}$ contains some $\mathfrak{a}_i$ by what we have shown, so must equal it in that $\mathfrak{p} \subseteq \mathfrak{a}$ by assumption. □

EXAMPLE 11.6.15. Let F be a field. The ideal (xy^2) of $F[x,y]$ has a minimal primary decomposition $(xy^2) = (x) \cap (y^2)$, and the associated primes of (xy^2) are (x) and (y^2).

THEOREM 11.6.16. *Let I be a decomposable ideal of R. The set of associated primes to the primary ideals in a minimal primary decomposition of I is uniquely determined by I.*

PROOF. Let $\{\mathfrak{q}_1, \dots, \mathfrak{q}_n\}$ be a primary decomposition of I. Now for $a \in R$, let $I_a = \{r \in R \mid ra \in I\}$, which is an ideal of R. We have $I_a = \bigcap_{i=1}^n (\mathfrak{q}_i)_a$. Let $\mathfrak{p}_i = \sqrt{\mathfrak{q}_i}$ for each i. Note that $(\mathfrak{q}_i)_a = R$ if $a \in \mathfrak{q}_i$, and $\sqrt{(\mathfrak{q}_i)_a} = \mathfrak{p}_i$ otherwise, so

$$\sqrt{I_a} = \bigcap_{i=1}^{n} \sqrt{(\mathfrak{q}_i)_a} = \bigcap_{\substack{i=1 \\ a \notin \mathfrak{q}_i}}^{n} \mathfrak{p}_i.$$

For any i, we may choose a in the intersection of the $\mathfrak{q}_j$ with $j \neq i$ such that $a_i \notin \mathfrak{q}_i$ by the minimality of our decomposition, and for such an a we have $\sqrt{I_a} = \mathfrak{p}_i$. On the other hand, for any $a \in R$ such that $\sqrt{I_a}$ is a prime ideal, part b of Lemma 11.6.14 tells us that $\sqrt{I_a} = \mathfrak{p}_i$ for some i (with $a \notin \mathfrak{q}_i$). Thus, the set of associated primes $\{\mathfrak{p}_1, \dots, \mathfrak{p}_n\}$ is uniquely determined by I. □

DEFINITION 11.6.17. Let I be a decomposable ideal of R. A prime ideal is called an *associated prime* of I if it is the associated prime of an element of a minimal primary decomposition of I.

DEFINITION 11.6.18. Let I be a proper ideal of R. An *isolated prime* of I is a minimal element in the set of prime ideals of R containing I, ordered by inclusion.

PROPOSITION 11.6.19. *Let I be a decomposable ideal of R. A prime ideal $\mathfrak{p}$ of R is an isolated prime of I if and only if it is a minimal element under inclusion in the set of associated primes of I.*

PROOF. Let $I = \bigcap_{i=1}^n \mathfrak{q}_i$ with each $\mathfrak{q}_i$ primary, and let $\mathfrak{p}_i$ be the associated prime of $\mathfrak{q}_i$. If $\mathfrak{p}$ is a prime ideal of R containing I, then

$$\mathfrak{p} = \sqrt{\mathfrak{p}} \supseteq \sqrt{I} = \bigcap_{i=1}^{n} \sqrt{\mathfrak{q}_i} = \bigcap_{i=1}^{n} \mathfrak{p}_i.$$

By part b of Lemma 11.6.14, we have that $\mathfrak{p}$ contains some $\mathfrak{p}_i$ for some i, so it contains some minimal prime in the set of associated primes of I. □

EXAMPLE 11.6.20. Let F be a field. The ideal $I = (xy, y^2)$ of $F[x,y]$ has a minimal primary decomposition $I = (x,y)^2 \cap (y)$, so it has associated primes (x,y) and (y). The ideal (y) is the unique isolated prime of I. Note that I also has the primary decomposition $I = (x, y^2) \cap (y)$.

In fact, the following uniqueness result also holds.

PROPOSITION 11.6.21. *Let I be a decomposable ideal of R. Let $\mathfrak{p}_1, \dots, \mathfrak{p}_n$ be distinct isolated primes of I. Let Q be a minimal primary decomposition of I, and let $\mathfrak{q}_i \in Q$ be $\mathfrak{p}_i$-primary for each $1 \leq i \leq n$. Then the ideal $\bigcap_{i=1}^n \mathfrak{q}_i$ is independent of the choice of Q.*

PROOF. We merely sketch the proof. Consider $S = R - \bigcup_{i=1}^{n} \mathfrak{p}_i$. For any associated prime ideal $\mathfrak{p}$ of I, the intersection $S \cap \mathfrak{p}$ is nonempty if and only if $\mathfrak{p} = \mathfrak{p}_i$ for some i by Proposition 11.1.14. Given a primary decomposition Q, suppose that $\mathfrak{q} \in Q$ is $\mathfrak{p}$-primary for some prime $\mathfrak{p}$. We have that $S^{-1}\mathfrak{p}_i$ is a prime of $S^{-1}R$ and $S^{-1}\mathfrak{q}_i$ is $S^{-1}\mathfrak{p}_i$-primary. For any other associated prime $\mathfrak{p}$ of R, we have that $S^{-1}\mathfrak{p} = S^{-1}R$. We then have

$$S^{-1}I = \bigcap_{\mathfrak{q} \in Q} S^{-1}\mathfrak{q} = \bigcap_{i=1}^{n} S^{-1}\mathfrak{q}_i,$$

and the contraction of $S^{-1}I$ to R is $\bigcap_{i=1}^{n} \mathfrak{q}_i$. Hence, the latter intersection is independent of the choice of Q. □

COROLLARY 11.6.22. *If $\mathfrak{p}$ is any isolated prime of a decomposable ideal I, then the unique $\mathfrak{p}$-primary ideal in any minimal primary decomposition of I is independent of the choice of decomposition.*

The reader may now easily check the following.

COROLLARY 11.6.23. *The primary ideals in a noetherian ring R are exactly the irreducible ideals.*

11.7. Hilbert's Nullstellensatz

We use K to denote a fixed algebraically closed field in this section. Much but certainly not all of what is done here can be generalized to fields which are not algebraically closed as well, but for this brief introduction, we feel it suffices to focus on the more specific setting. This section assumes some basic knowledge of topological spaces.

Fix a nonnegative integer n.

DEFINITION 11.7.1. Let S be a subset of $K[x_1,\dots,x_n]$. The *zero set*, or *vanishing locus*, of S is

$$V(S) = \{(a_1,\dots,a_n) \in K^n \mid f(a_1,\dots,a_n) = 0 \text{ for all } f \in S\}.$$

An *algebraic set* in K^n is any subset of K^n that is a zero set of some set of polynomials in $K[x_1,\dots,x_n]$.

From now on, let us set $R = K[x_1,\dots,x_n]$ for brevity.

NOTATION 11.7.2. If $S = \{f_1,\dots,f_n\} \subset R$, we also write $V(f_1,\dots,f_n)$ for $V(S)$. At times, for $a = (a_1,\dots,a_n) \in K^n$, we write $f(a)$ for $f(a_1,\dots,a_n)$.

EXAMPLE 11.7.3. We have $V(\varnothing) = X$ and $V(R) = V(1) = \varnothing$.

EXAMPLE 11.7.4. Consider $f(x,y) = x - y$ and $g(x,y) = x^2 + y^2 - 2$ in $\mathbb{C}[x,y]$. Then $V(f,g) = V(f) \cap V(g) = \{(-1,-1),(1,1)\}$.

REMARK 11.7.5. For any subset S of R, the zero set $V(S)$ equals the zero set of the ideal (S) generated by S.

PROPOSITION 11.7.6.

a. The intersection of any collection of algebraic sets in K^n is also an algebraic set.

b. The union of any finite collection of algebraic sets in K^n is also an algebraic set.

PROOF. Let $\{S_i \mid i \in I\}$ be a collection of subsets of R. Then $\bigcap_{i\in I} V(S_i) = V(S)$ is algebraic, so we have part a. If S and T are subsets of R, set $I = (S)$ and $J = (T)$. We clearly have

$$V(S) \cup V(T) = V(I) \cup V(J) \subseteq V(I \cap J).$$

If $a \in V(I \cap J)$ and $a \notin V(I)$, then there exists $f \in I$ with $f(a) \neq 0$. If $g \in J$, then $fg \in I \cap J$, so $f(a)g(a) = 0$, so $g(a) = 0$. Thus $a \in V(J)$, and we have part b. □

It follows from the proposition that the following definition does in fact yield a topology.

DEFINITION 11.7.7. The *Zariski topology* on K^n is the topology $\{K^n - V(S) \mid S \subseteq R\}$ with closed sets the algebraic sets in K^n.

DEFINITION 11.7.8. For $n \geq 0$, the *affine n-space* over K is the set $\mathbb{A}^n_K = K^n$ endowed with the Zariski topology.

REMARK 11.7.9. For any $(a_1, \ldots, a_n) \in \mathbb{A}^n_K$, we have $V(x_1 - a_1, \ldots, x_n - a_n) = \{(a_1, \ldots, a_n)\}$, so points in $\mathbb{A}^n_K$ are closed. However, it is not a Hausdoff topology: for instance, for $n = 1$, the only closed sets other than X are finite, so any two nonempty open sets will intersect as K is infinite.

NOTATION 11.7.10. Let $Z \subseteq \mathbb{A}^n_K$. Then

$$I(Z) = \{f \in K[x_1, \ldots, x_n] \mid f(a) = 0 \text{ for all } a \in Z\}.$$

The set $I(Z)$ is clearly an ideal: it is the ideal of R of elements that vanish on all of Z.

REMARK 11.7.11. Note that if $f \in R$ satisfies $f^k \in I(Z)$ for some subset Z of $\mathbb{A}^n_K$ and $k \geq 1$, then $f(a)^k = 0$ for all $a \in Z$, so f vanishes on Z, which is to say that $f \in I(Z)$. Hence, $I(Z)$ is a radical ideal.

EXAMPLE 11.7.12. For $a = (a_1, \ldots, a_n) \in \mathbb{A}^n_K$, we have $I(\{a\}) = (x_1 - a_1, \ldots, x_n - a_n)$.

In particular I and V provide bijections between the points of $\mathbb{A}^n_K$ and a subset of the maximal ideals of R, i.e., those of the form $(x_1 - a_1, \ldots, x_n - a_n)$ for some $a \in \mathbb{A}^n_K$. The statement the latter maximal ideals are all of the maximal ideals of R is known as the weak form of Hilbert's Nullstellensatz.

THEOREM 11.7.13. *Every maximal ideal of $K[x_1, \ldots, x_n]$ has the form $(x_1 - a_1, \ldots, x_n - a_n)$ for some $(a_1, \ldots, a_n) \in \mathbb{A}^n_K$.*

PROOF. Let $\mathfrak{m}$ be a maximal ideal of $R = K[x_1, \ldots, x_n]$, and consider $L = R/\mathfrak{m}$, which is a field containing K that is finitely generated over K. By Corollary 11.3.44, the field L is an algebraic extension of the algebraically closed field K, so it is equal to K. Under the quotient map $R \to L = K$, each x_i is sent to some $a_i \in L$, so $x_i - a_i \in \mathfrak{m}$. Since $(x_1 - a_1, \ldots, x_n - a_n)$ is maximal, it equals $\mathfrak{m}$. □

In other words, V and I give inverse bijections between the maximal ideals of R and the singleton subsets of $\mathbb{A}^n_K$. We now prove the stronger form of this statement, one which boils down to the statement that $I(V(\mathfrak{a})) = \sqrt{\mathfrak{a}}$ for ideals $\mathfrak{a}$ of R.

THEOREM 11.7.14 (Hilbert's Nullstellensatz). *The maps I and V provide mutually inverse, inclusion-reversing bijections*

$$\{\textit{radical ideals of } K[x_1,\ldots,x_n]\} \underset{I}{\overset{V}{\rightleftarrows}} \{\textit{algebraic sets in } \mathbb{A}^n_K\}.$$

PROOF. The operation I is by definition inclusion-reversing on subsets of $\mathbb{A}^n_K$, and the operation V is inclusion-reversing on subsets of $R = K[x_1,\ldots,x_n]$. It is immediate from the definitions and Remark 11.7.11 that if $\mathfrak{a}$ is an ideal of R, then $I(V(\mathfrak{a}))$ contains $\sqrt{\mathfrak{a}}$, and if Z is a subset of R, then $V(I(Z))$ contains Z. If $Z = V(\mathfrak{a})$ for some ideal $\mathfrak{a}$, then

$$V(I(Z)) = V(I(V(\mathfrak{a}))) \subseteq V(\mathfrak{a}) = Z,$$

since V is inclusion-reversing. Thus, on algebraic sets Z, we have $V(I(Z)) = Z$.

It remains to show that $I(V(\mathfrak{a})) \subseteq \sqrt{\mathfrak{a}}$ for any ideal $\mathfrak{a}$ of R. Let $f \in I(V(\mathfrak{a}))$. Since R is noetherian, we have $\mathfrak{a} = (g_1,\ldots,g_k)$ for some $g_1,\ldots,g_k \in R$. For an indeterminate y, let J be the ideal of $R[y]$ generated by I and $1 - fy$. We view $R[y]$ as $K[x_1,\ldots,x_n,y]$ and consider the vanishing set of J in $\mathbb{A}^{n+1}_K$. If $a = (a_1,\ldots,a_{n+1}) \in V(J)$, then $(a_1,\ldots,a_n) \in V(\mathfrak{a})$, in which case we have

$$(1 - fy)(a_1,\ldots,a_{n+1}) = 1 - f(a_1,\ldots,a_n)a_{n+1} = 1 \neq 0.$$

Thus $V(J) = \varnothing$. By the weak form of the Nullstellensatz, if J were a proper ideal, then its vanishing locus would contain the point in the vanishing locus of a maximal ideal containing it, so $J = R[y]$.

Since $J = R[y]$, we may write

$$1 = h(1 - fy) + \sum_{i=1}^{k} h_i g_i$$

with $h \in R[y]$ and $h_i \in R[y]$ for $1 \leq i \leq k$. Let N be the maximum of the degree of the h_i's with $1 \leq i \leq k$ and hy as polynomials in y. Set $z = y^{-1}$. Multiplying our equation for 1 by z^{N+1}, we have

$$z^{N+1} = h'(y - f) + \sum_{i=1}^{k} h_i' g_i$$

for some $h' \in R[z]$ and $h_i' \in R[z]$. Substituting in $z = f$, we then have $f^{N+1} \in (g_1,\ldots,g_k) = \mathfrak{a}$. That is, $f \in \sqrt{\mathfrak{a}}$, as was desired. □

REMARK 11.7.15. We record the following simple consequences of the Nullstellensatz.

a. For any $S \subseteq R$, we have $V(S) = V(\sqrt{(S)})$, and $I(V(S)) = \sqrt{(S)}$.

b. For any $Z \subseteq \mathbb{A}^n_K$, we have $I(Z) = I(\overline{Z})$, where $\overline{Z}$ is the closure of Z in the Zariski topology (i.e., the smallest algebraic set containing Z), and $V(I(Z)) = \overline{Z}$.

DEFINITION 11.7.16. We say that an algebraic set is *irreducible* if it is not a union of two proper algebraic subsets.

We have seen that maximal ideals correspond to singleton sets under V and I. Hilbert's Nullstellensatz tells us that prime ideals correspond to irreducible algebraic sets.

COROLLARY 11.7.17. *The maps I and V restrict to mutually inverse, inclusion-reversing bijections*

$$\{\textit{prime ideals of } K[x_1,\dots,x_n]\} \underset{I}{\overset{V}{\rightleftarrows}} \{\textit{irreducible algebraic sets in } \mathbb{A}^n_K\}.$$

PROOF. An algebraic set Z is by definition the vanishing locus of some radical ideal I of R. By the Nullstellensatz, such a set Z is irreducible if and only if $I = I(Z)$ cannot be written as an intersection of two radical ideals properly containing I. Note that if I were the intersection of two arbitrary ideals, then it would also be the intersection of their radicals. Conversely, if I is an irreducible ideal, then so is its radical, and then its vanishing locus is irreducible as well. Since the irreducible ideals in R are exactly the primary ideals, and those which are radical are the prime ideals, irreducible algebraic sets correspond exactly to the prime ideals of R. □

REMARK 11.7.18. By the primary decomposition theorem and Corollary 11.7.17, every algebraic set is a finite union of irreducible algebraic sets.

REMARK 11.7.19. An irreducible algebraic set $Z \subseteq \mathbb{A}^n_K$ together with its subspace topology is also what is called an (affine) algebraic variety. It has an associated coordinate ring $K[Z] = R/I(Z)$. Note that the ring $R/I(Z)$ is reduced, i.e., has no nilpotents, since $I(Z)$ is a radical ideal. The radical ideals of $K[Z]$ correspond to algebraic subsets of Z, and via this bijection the maximal ideals of $K[Z]$ correspond to the points (or more precisely, singleton subsets) of Z.

11.8. Spectra of rings

In the previous section, we say that for the points of $\mathbb{A}^n_K$ for an algebraically closed field K correspond to the maximal ideals of $K[x_1,\dots,x_n]$. Making this identification, we may think of the Zariski topology as endowing the set of maximal ideals of $K[x_1,\dots,x_n]$ with a topology. We now aim to mimic this for the larger set of prime ideals, in an arbitrary commutative ring R.

DEFINITION 11.8.1. The *spectrum* $\operatorname{Spec} R$ of a commutative ring R is the set of prime ideals of R.

EXAMPLE 11.8.2. For a PID R, we have $\operatorname{Spec} R = \{(0)\} \cup \{(f) \mid f \text{ irreducible}\}$.

NOTATION 11.8.3. For any subset T of R, we set

$$V(T) = \{\mathfrak{p} \in \operatorname{Spec} R \mid T \subseteq \mathfrak{p}\}.$$

For any subset Y of $\operatorname{Spec} R$, we set

$$I(Y) = \bigcap_{\mathfrak{p} \in Y} \mathfrak{p}.$$

REMARK 11.8.4. We have $V(T) = V((T))$ for the ideal (T) generated by T. In fact, for any ideal I of R, we have $V(I) = V(\sqrt{I})$ since if $I \subseteq \mathfrak{p}$, then $\sqrt{I} \subseteq \sqrt{\mathfrak{p}} = \mathfrak{p}$.

The following lemma is easily verified.

LEMMA 11.8.5.

a. We have $V((0)) = \operatorname{Spec} R$ *and* $V(R) = \varnothing$.

b. If $\mathfrak{a}$ *and* $\mathfrak{b}$ *are ideals of* R*, then* $V(\mathfrak{a}) \cup V(\mathfrak{b}) = V(\mathfrak{a} \cap \mathfrak{b}) = V(\mathfrak{a}\mathfrak{b})$.

c. If $\{\mathfrak{a}_j \mid j \in X\}$ *is a collection of ideals of* R*, then*

$$\bigcap_{j \in X} V(\mathfrak{a}_j) = V\left(\bigcup_{j \in X} \mathfrak{a}_j\right) = V\left(\sum_{j \in X} \mathfrak{a}_j\right).$$

In particular, the sets $V(I)$ for I an ideal of R form a topology on $\operatorname{Spec} R$.

DEFINITION 11.8.6. The *Zariski topology* on $\operatorname{Spec} R$ is the unique topology with closed sets the $V(I)$ with I an ideal of R.

REMARK 11.8.7. In $\operatorname{Spec} R$, the singleton sets $\{\mathfrak{m}\}$ with $\mathfrak{m}$ maximal are closed, since $V(\mathfrak{m}) = \{\mathfrak{m}\}$. However, points in general need not be closed. The closure of $\{\mathfrak{p}\}$ with $\mathfrak{p}$ prime is the smallest closed subset containing $\mathfrak{p}$, which is exactly $V(\mathfrak{p})$, the set of prime ideals containing $\mathfrak{p}$. So, $\{\mathfrak{p}\}$ is closed if and only if $\mathfrak{p}$ is maximal. E.g., in an integral domain, the closure of (0) is $\operatorname{Spec} R$!

DEFINITION 11.8.8. The *closed points* of $\operatorname{Spec} R$ are the maximal ideals of R.

DEFINITION 11.8.9. The closure of a subset Y of $\operatorname{Spec} R$ in the Zariski topology on R is known as the *Zariski closure* of Y.

The analogue of the Nullstellensatz for $\operatorname{Spec} R$ is considerably less difficult.

PROPOSITION 11.8.10. *The maps I and V provide mutually inverse, inclusion-reversing bijections*

$$\{\textit{radical ideals of } R\} \underset{I}{\overset{V}{\rightleftarrows}} \{\textit{closed subsets of } \operatorname{Spec} R\}.$$

In fact, for any ideal $\mathfrak{a}$ *of* R*, we have* $I(V(\mathfrak{a})) = \sqrt{\mathfrak{a}}$*, and for any subset* Y *of* $\operatorname{Spec} R$*, the set* $V(I(Y))$ *is the Zariski closure of* Y.

PROOF. That V and I are inclusion-reversing is clear. Let $\mathfrak{a}$ be an ideal of R. Then

$$I(V(\mathfrak{a})) = \bigcap_{\mathfrak{a} \subseteq \mathfrak{p}} \mathfrak{p} = \sqrt{\mathfrak{a}}$$

by Proposition 11.4.7. Conversely, if Y is a subset of $\operatorname{Spec} R$, then its closure $\overline{Y}$ is $V(\mathfrak{a})$ for some ideal $\mathfrak{a}$ of R, and

$$V(I(Y)) \subseteq V(I(\overline{Y})) = V(I(V(\mathfrak{a}))) = V(\sqrt{\mathfrak{a}}) = V(\mathfrak{a}) = \overline{Y},$$

but $V(I(Y))$ is closed an contains Y, so $V(I(Y)) = \overline{Y}$. □

COROLLARY 11.8.11. *The Zariski closure of a subset Y of* $\operatorname{Spec} R$ *is the set of all prime ideals containing some element of Y.*

PROOF. The set $V(I(Y))$ consists of the prime ideals $\mathfrak{p}$ containing the intersection of all prime ideals containing Y. By Lemma 11.6.14b, these are exactly the ideals that contain some element of Y. □

Let us compare V and I of Definition 11.8.3 with our prior maps with these notations for the polynomial ring $R = K[x_1,\dots,x_n]$ over an algebraically closed field K.

PROPOSITION 11.8.12. *Let $R = K[x_1,\dots,x_n]$ for some $n \geq 0$ and algebraically closed field K. Let us use V' and I' to denote the maps which take vanishing loci of algebraic sets in $\mathbb{A}^n_K$ and the ideal of vanishing of subsets of R, respectively.*

a. The injective map $\iota\colon \mathbb{A}^n_K \to \operatorname{Spec} R$ given by taking a point to its corresponding maximal ideal is a homeomorphism onto its image, which we use to identify $\mathbb{A}^n_K$ with a subspace of $\operatorname{Spec} R$.

b. For any ideal $\mathfrak{a}$ of R, we have $V'(\mathfrak{a}) = V(\mathfrak{a}) \cap \mathbb{A}^n_K$.

c. For any Zariski closed subset Y of $\operatorname{Spec} R$*, we have $I(Y) = I'(Y \cap \mathbb{A}^n_K)$.*

PROOF. If $Y = V(\mathfrak{a})$ for some ideal $\mathfrak{a}$ of $\operatorname{Spec} R$, then $Y \cap \mathbb{A}^n_K$ is the set of maximal ideals of R containing $\mathfrak{a}$, which equals $V'(\mathfrak{a})$, proving part a. This implies that the intersection of Y with $\mathbb{A}^n_K$ is closed and that the image of a closed set $Z = V'(\mathfrak{a})$ in the Zariski topology on $\mathbb{A}^n_K$ is closed under the subspace topology on $\mathfrak{A}^n_K$ from the Zariski topology on $\operatorname{Spec} R$. Thus, ι is a homeomorphism onto its image, proving part b.

Finally, if $Y = V(\mathfrak{a})$ is a closed subset of $\operatorname{Spec} R$ with $\mathfrak{a}$ radical, and if $Z = Y \cap \mathbb{A}^n_K = V'(\mathfrak{a})$, then $\mathfrak{a} = I'(Z)$ is the intersection of all prime ideals containing $\mathfrak{a}$. That is, $\mathfrak{a}$ is the intersection of all prime ideals in $V(\mathfrak{a}) = Y$, so $\mathfrak{a} = I(Y)$ as well, and we have part c. □

DEFINITION 11.8.13. If $a \in R$, then $U_a = \operatorname{Spec} R - V((a))$ is called a *principal open set* of R.

PROPOSITION 11.8.14. *The sets U_a for $a \in R$ form a basis for the Zariski topology on* $\operatorname{Spec} R$.

PROOF. Let U be an open set in $\operatorname{Spec} R$. Then $U = \operatorname{Spec} R - V(I)$ for some ideal I, and $V(I) = \bigcap_{a \in I} V((a))$, so $U = \bigcup_{a \in I} U_a$. □

We have the following simple lemma.

LEMMA 11.8.15. *Let $\varphi\colon R \to S$ be a ring homomorphism.*

a. If $\mathfrak{q}$ is a prime ideal of S, then $\varphi^{-1}(\mathfrak{q})$ is a prime ideal of R.

b. Suppose that φ is surjective and $\mathfrak{p}$ is a prime ideal of R containing the kernel of φ. Then $\varphi(\mathfrak{p})$ is a prime ideal of S.

PROOF. Set $\mathfrak{p} = \varphi^{-1}(\mathfrak{q})$ for $\mathfrak{q}$ a prime ideal of S. Then $a,b \in R$ satisfy $ab \in \mathfrak{p}$ if and only if $\varphi(ab) \in \mathfrak{q}$, so if and only if either $\varphi(a) \in \mathfrak{q}$ or $\varphi(b) \in \mathfrak{q}$, i.e., $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$.

If $\mathfrak{p}$ is a prime ideal of R containing $\ker\varphi$, then $\mathfrak{q} = \varphi(\mathfrak{p})$ is an ideal of S by the surjectivity of φ, For $a,b \in S$ with $ab \in \mathfrak{q}$, write $a = \varphi(c)$ and $b = \varphi(d)$ for some $c,d \in R$. Then $cd \in \varphi^{-1}(\mathfrak{q}) = \mathfrak{p}$ since $\ker\varphi \subseteq \mathfrak{p}$, so $c \in \mathfrak{p}$ or $d \in \mathfrak{p}$, and therefore $a \in \mathfrak{q}$ or $b \in \mathfrak{q}$. □

By Lemma 11.8.15, the following definition makes sense.

DEFINITION 11.8.16. Let $\varphi\colon R \to S$ be a ring homomorphism. The *pullback map*

$$\varphi^*\colon \operatorname{Spec} S \to \operatorname{Spec} R$$

is the function given by $\varphi^*(\mathfrak{q}) = \varphi^{-1}(\mathfrak{q})$ for $\mathfrak{q} \in \operatorname{Spec} S$.

The following lemma is simple.

LEMMA 11.8.17. *If φ is a ring homomorphism, then the pullback map φ^* is continuous with respect to the Zariski topologies.*

REMARK 11.8.18. The map that takes a ring to its spectrum and ring homomorphism to the corresponding pullback map is a contravariant functor from **Ring** to **Top**.

EXAMPLE 11.8.19. Let $\pi\colon R \to F$ be a surjective ring homomorphism, where F is a field. Then $\pi^*((0)) = \ker\pi$ is the maximal ideal that is the kernel of f.

EXAMPLE 11.8.20. Let $\mathfrak{p}$ be a prime ideal. The localization map $\varphi\colon R \to R_{\mathfrak{p}}$ has pullback $\varphi^*(\mathfrak{q}R_{\mathfrak{p}}) = \mathfrak{q}$ for prime ideals $\mathfrak{q}$ of R contained in $\mathfrak{p}$.

EXAMPLE 11.8.21. Consider the map $\varphi\colon \mathbb{C}[x] \to \mathbb{C}[x]$ given by $\varphi(f)(x) = f(x^2)$. Then $\varphi^*((0)) = (0)$, and for $a \in \mathbb{C}$ irreducible,

$$\varphi^*((x-a)) = \{g \in K[x] \mid g(a^2) = 0\} = (x-a^2).$$

In particular, φ^* is 2-to-1, taking both $(x-a)$ and $(x+a)$ to $(x-a^2)$, except for $a = 0$, in which case only (x) is carried to (x).

11.9. Krull dimension

We continue to use R to denote a commutative ring.

DEFINITION 11.9.1. The *length* of an ascending chain $(\mathfrak{p}_i)_{i=0}^n$ of distinct prime ideals is n. We often refer to such a finite strictly ascending chain more simply as a *chain of prime ideals*, where minimal confusion can arise.

EXAMPLE 11.9.2. If R is an integral domain and $n \geq 0$, then the ring $R[x_1,\ldots,x_n]$ contains a chain of primes of length n:

$$(0) \subset (x_1) \subset (x_1,x_2) \subset \cdots \subset (x_1,\ldots,x_n).$$

DEFINITION 11.9.3. The *Krull dimension*, or *dimension*, $\dim R$ of a commutative ring R is the length of the longest ascending chain of distinct prime ideals in R, if it exists, and is otherwise said to be infinite.

REMARK 11.9.4. In set-theoretic terms, if finite, $\dim R$ is one less than the maximum of the cardinalities of all chains in $\operatorname{Spec} R$.

EXAMPLES 11.9.5. Let F be a field.

a. The Krull dimension of F is 0: its only prime ideal is (0). In fact, the Krull dimension of $F[x]/(x^n)$ for $n \geq 0$ is 1, since its unique prime ideal is (x).

b. The Krull dimension of $\mathbb{Z}$ is 1: the longest chains are all of the form $(0) \subset (p)$ for some prime number p. In fact, $\dim R = 1$ for every PID R that is not a field.

c. The Krull dimension of $F[(x_i)_{i\geq 1}]$ is infinite, since

$$(0) \subset (x_1) \subset (x_1, x_2) \subset \cdots$$

is an ascending chain of prime ideals that is not eventually constant.

LEMMA 11.9.6. *Let $\pi\colon R \to S$ be a surjective map of rings. Then* $\dim R \geq \dim S$.

PROOF. Let $(\mathfrak{q}_i)_{i=0}^n$ be a chain of primes in S of length n, and set $\mathfrak{p}_i = \varphi^{-1}(\mathfrak{q}_i)$ for each i. Then $\mathfrak{q}_i = \varphi(\mathfrak{p}_i)$ for each i, so each $\mathfrak{p}_i$ is distinct. □

LEMMA 11.9.7. *Let $\mathfrak{p}$ be a non-minimal prime of R. Then* $\dim R \geq \dim R/\mathfrak{p} + 1$.

PROOF. Any chain in $R/\mathfrak{p}$ of maximal length has inverse image in R of the same length, and such a chain can be extended by adding in a minimal prime properly contained in $\mathfrak{p}$. □

PROPOSITION 11.9.8. *If B/A is an integral extension of domains, then A has finite Krull dimension if and only if B does, in which case* $\dim B = \dim A$.

PROOF. The the going up theorem tells us that $\dim B \geq \dim A$. Suppose that $\dim B = n$, let $(\mathfrak{q}_i)_{i=0}^n$ be a maximal ascending chain of prime ideals of B, and set $\mathfrak{p}_i = \mathfrak{q}_i \cap A$ for all i. We have $\mathfrak{p}_1 \cap A \neq (0)$ by Lemma 11.5.2, so $\dim B/\mathfrak{q}_1 = \dim B - 1$, and $\dim A/\mathfrak{p}_1 \leq \dim A - 1$ by Lemma 11.9.7. By induction on $\dim B$, we then have the remaining inequality $\dim B \leq \dim A$. □

PROPOSITION 11.9.9. *Let F be a field and n be a nonnegative integer. The ring $F[x_1, \dots, x_n]$ has Krull dimension n.*

PROOF. Example 11.9.2 tells us that $R = F[x_1, \dots, x_n]$ has dimension at least n. We may suppose that $n \geq 1$. Let $(\mathfrak{p}_i)_{i=0}^m$ be an ascending chain of prime ideals in R. We may suppose that $\mathfrak{p}_0 = (0)$ and that $\mathfrak{p}_1$ is minimal, generated by an irreducible element $g \in R$, as otherwise we may extend the chain to contain such primes.

Consider the quotient $R = S/\mathfrak{p}_1 = S/(g)$. Since the images of the x_i in S satisfy an equation of algebraic dependence over F, and these images generate S as an F-algebra, the quotient field of S has transcendence degree at most $n-1$ over F. Thus, no set of more than $n-1$ elements of S can be algebraically independent.

By Noether's normalization lemma, there exist algebraically independent elements $t_1, \dots, t_s \in R$ such that S is integral over $F[t_1, \dots, t_s]$. From what we have already shown, we must have $s \leq n-1$. Then $\dim S = s \leq n-1$ by Proposition 11.9.8 and induction. On the other hand, the images $\bar{\mathfrak{p}}_i$ in S of the ideals $\mathfrak{p}_i$ with $1 \leq i \leq m$ remain prime in S by Lemma 11.8.15, and they are distinct, so $m - 1 \leq s \leq n-1$. Therefore, $m = n$. □

In fact, the following result, for which we omit the proof, holds more generally.

THEOREM 11.9.10. *Let R be a noetherian domain of finite Krull dimension. Then*

$$\dim R[x] = \dim R + 1.$$

DEFINITION 11.9.11. The *height* $\mathrm{ht}(\mathfrak{p})$ of a prime ideal $\mathfrak{p}$ of R is the length of the longest chain of primes of R contained in $\mathfrak{p}$. A prime of height 0 is called a *minimal prime* of R.

Note that a minimal prime of R is just an isolated prime of (0).

EXAMPLES 11.9.12.

a. In $F[x_1,\dots,x_n]$ for a field F, the height of $(x_1,\dots,x_k)$ for $k \le n$ is k.

b. In a UFD, the primes of height one are principal, generated by the irreducible elements.

c. In a product of fields $R = \prod_{i=1}^n F_i$, the minimal primes are the maximal ideals, the kernels of projection maps $R \to F_k$ for some $1 \le k \le n$.

REMARK 11.9.13. Suppose $R = K[x_1,\dots,x_n]$ with K algebraically closed. The prime ideals $\mathfrak{p}$ of R correspond to algebraic sets in $\mathbb{A}_K^n$. The dimension of the algebraic set V that is the vanishing locus of $\mathfrak{p}$ is defined to be $n - \operatorname{ht}(\mathfrak{p})$. In particular, $\mathbb{A}_K^n$ has dimension n, as one would expect. We often refer to $\operatorname{ht}(\mathfrak{p})$ as the codimension of V in $\mathbb{A}_K^n$. In particular, the vanishing locus of a single nonconstant polynomial in R has codimension 1.

11.10. Dedekind domains

DEFINITION 11.10.1. A *Dedekind domain* is a noetherian, integrally closed domain of Krull dimension at most 1.

The condition of having Krull dimension at most 1 is the same as every nonzero prime ideal being maximal. We have the following class of examples.

LEMMA 11.10.2. *Every PID is a Dedekind domain.*

PROOF. A PID is noetherian, and it is a UFD, so it is integrally closed. Its nonzero prime ideals are maximal, generated by its irreducible elements. □

COROLLARY 11.10.3. *Let A be a Dedekind domain, and let B be the integral closure of A in a finite, separable extension of the quotient field of A. Then B is a Dedekind domain.*

PROOF. Note that B is a finitely generated A-module by Corollary 11.3.37. If $\mathfrak{b}$ is an ideal of B, then $\mathfrak{b}$ is an A-submodule of B, and as A is noetherian, it is therefore finitely generated. Thus, B is noetherian. That B is integrally closed is just Proposition 11.3.25. That every nonzero prime ideal in A is maximal follows from Lemma 11.9.8 □

We have the following immediate corollary.

COROLLARY 11.10.4. *The ring of integers of any number field is a Dedekind domain.*

More examples of Dedekind domains can be produced as follows.

PROPOSITION 11.10.5. *Let A be a Dedekind domain, and let S be a multiplicatively closed subset of A. Then $S^{-1}A$ is also a Dedekind domain.*

PROOF. Given an ideal $\mathfrak{b}$ of $S^{-1}A$, set $\mathfrak{a} = A \cap \mathfrak{b}$. Then $\mathfrak{a}$ is an ideal of A, and $\mathfrak{b} = S^{-1}\mathfrak{a}$. It follows that any set of generators of $\mathfrak{a}$ as an ideal of A generates $S^{-1}\mathfrak{a}$ as an ideal of $S^{-1}A$. Hence $S^{-1}A$ is noetherian. If, moreover, $\mathfrak{b}$ is a nonzero prime, then clearly $\mathfrak{a}$ is as well, and $\mathfrak{a}$ is maximal since A is a Dedekind domain. Then $S^{-1}A/\mathfrak{b} \cong A/\mathfrak{a}$ is a field, so $\mathfrak{b}$ is maximal as well.

Let K be the quotient field of A. Any $\alpha \in K$ that is integral over $S^{-1}A$ satisfies a monic polynomial f with coefficients in $S^{-1}A$. Set $n = \deg f$. If $d \in S$ is the product of the denominators of these coefficients, then $d^n f(d^{-1}x) \in A[x]$ is monic with $d\alpha \in K$ as a root. Since A is integrally closed, we have $d\alpha \in A$, so $\alpha \in S^{-1}A$. That is, $S^{-1}A$ is integrally closed. □

LEMMA 11.10.6. *Let A be a noetherian domain, and let $\mathfrak{a}$ be a nonzero ideal of A.*

a. There exist $k \geq 0$ and nonzero prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_k$ of A such that $\mathfrak{p}_1 \cdots \mathfrak{p}_k \subseteq \mathfrak{a}$.

b. Suppose that $\dim A \leq 1$. *If $\mathfrak{p}_1, \ldots, \mathfrak{p}_k$ are as in part a and $\mathfrak{p}$ is a prime ideal of A containing $\mathfrak{a}$, then $\mathfrak{p} = \mathfrak{p}_i$ for some positive $i \leq k$.*

PROOF. Consider the set X of nonzero ideals of A for which the statement of the first part of the lemma fails, and order X by inclusion. Suppose by way of contradiction that X is nonempty. Let C be a chain in X. Either C has a maximal element or there exist $\mathfrak{a}_i \in C$ for $i \geq 1$ with $\mathfrak{a}_i \subsetneq \mathfrak{a}_{i+1}$ for each i. The latter is impossible as A is a noetherian. By Zorn's lemma, X contains a maximal element $\mathfrak{a}$. Now $\mathfrak{a}$ is not prime since it lies in X, so let $a, b \in A - \mathfrak{a}$ with $ab \in \mathfrak{a}$. Then $\mathfrak{a} + (a)$ and $\mathfrak{a} + (b)$ both properly contain $\mathfrak{a}$, so by maximality of $\mathfrak{a}$, there exist prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_k$ and $\mathfrak{q}_1, \ldots, \mathfrak{q}_l$ of A for some $k, l \geq 0$ such that $\mathfrak{p}_1 \cdots \mathfrak{p}_k \subseteq \mathfrak{a} + (a)$ and $\mathfrak{q}_1 \cdots \mathfrak{q}_l \subseteq \mathfrak{a} + (b)$. We then have

$$\mathfrak{p}_1 \cdots \mathfrak{p}_k \mathfrak{q}_1 \cdots \mathfrak{q}_l \subseteq (\mathfrak{a} + (a))(\mathfrak{a} + (b)) \subseteq \mathfrak{a},$$

a contradiction of $\mathfrak{a} \in X$. This proves part a.

Now, suppose that $\mathfrak{a}$ is proper, and let $\mathfrak{p}$ be a prime ideal containing $\mathfrak{a}$. Assume that $\dim A \leq 1$. If no $\mathfrak{p}_i$ equals $\mathfrak{p}$, then since $\mathfrak{p}_i$ is maximal, there exist $b_i \in \mathfrak{p}_i - \mathfrak{p}$ for each $1 \leq i \leq k$. We then have $b_1 \cdots b_k \notin \mathfrak{p}$ as $\mathfrak{p}$ is prime, so $b_1 \cdots b_k \notin \mathfrak{a}$, a contradiction. Hence we have part b. □

DEFINITION 11.10.7. A *fractional ideal* of a domain A is a nonzero A-submodule $\mathfrak{a}$ of the quotient field of A for which there exists a nonzero $d \in A$ such that $d\mathfrak{a} \subseteq A$.

REMARK 11.10.8. Every nonzero ideal in a domain A is a fractional ideal, which is sometimes referred to as an integral ideal. Every fractional ideal of A that is contained in A is an integral ideal.

EXAMPLE 11.10.9. The fractional ideals of $\mathbb{Z}$ are exactly the $\mathbb{Z}$-submodules of $\mathbb{Q}$ generated by a nonzero rational number.

LEMMA 11.10.10. *Let A be a noetherian domain. An A-submodule of the quotient field of A is a fractional ideal if and only if it is finitely generated.*

PROOF. If $\mathfrak{a}$ is a finitely generated A-submodule of the quotient field of A, then let $d \in A$ denote the product of the denominators of a set of generators. Then $d\mathfrak{a} \subseteq A$. Conversely, suppose that $\mathfrak{a}$ is a fractional ideal and $d \in A$ is nonzero and satisfies $d\mathfrak{a} \subseteq A$. Then $d\mathfrak{a}$ is an ideal of A, hence finitely generated. Moreover, the multiplication-by-d map carries $\mathfrak{a}$ isomorphically onto $d\mathfrak{a}$. □

DEFINITION 11.10.11. Let A be a domain with quotient field K, and let $\mathfrak{a}$ and $\mathfrak{b}$ be fractional ideals of A.

a. The inverse of $\mathfrak{a}$ is $\mathfrak{a}^{-1} = \{b \in K \mid b\mathfrak{a} \subseteq A\}$.

b. The product of $\mathfrak{a}$ and $\mathfrak{b}$ is the A-submodule of K generated by the set $\{ab \mid a \in \mathfrak{a}, b \in \mathfrak{b}\}$.

LEMMA 11.10.12. *Let A be a domain, and let $\mathfrak{a}$ and $\mathfrak{b}$ be fractional ideals of A. Then $\mathfrak{a}^{-1}$, $\mathfrak{a}+\mathfrak{b}$, $\mathfrak{a}\mathfrak{b}$, $\mathfrak{a}\cap\mathfrak{b}$ are fractional ideals of A as well.*

PROOF. Let K denote the quotient field of A. Let $c, d \in A$ be nonzero such that $c\mathfrak{a} \subseteq A$ and $d\mathfrak{b} \subseteq A$. Then $c(\mathfrak{a}\cap\mathfrak{b}) \subseteq A$, $cd(\mathfrak{a}+\mathfrak{b}) \subseteq A$, and $cd\mathfrak{a}\mathfrak{b} \subseteq A$.

Note that $\mathfrak{a}^{-1}$ is an A-submodule of K which is nonzero since there exists $d \in A$ with $d\mathfrak{a} \subseteq A$ in that $\mathfrak{a}$ is a fractional ideal. Let $a \in \mathfrak{a}$ be nonzero, and let $e \in A$ be its numerator in a representation of a as a fraction, so $e \in \mathfrak{a}$ as well. For any $t \in \mathfrak{a}^{-1}$, we have $te \in A$ by definition, so $e\mathfrak{a}^{-1} \subseteq A$, and therefore $\mathfrak{a}^{-1}$ is a fractional ideal. □

REMARK 11.10.13. By definition, multiplication of fractional ideals is an associative (and commutative) operation, so the set $I(A)$ of fractional ideals in A is a monoid.

DEFINITION 11.10.14. We say that a fractional ideal $\mathfrak{a}$ of a domain A is *invertible* if there exists a fractional ideal $\mathfrak{b}$ of A such that $\mathfrak{a}\mathfrak{b} = A$.

LEMMA 11.10.15. *A fractional ideal $\mathfrak{a}$ of a domain A is invertible if and only if $\mathfrak{a}^{-1}\mathfrak{a} = A$.*

PROOF. For the nonobvious direction, suppose that $\mathfrak{a}$ is invertible. Then we must have $\mathfrak{b} \subseteq \mathfrak{a}^{-1}$ by definition of $\mathfrak{a}^{-1}$. On the other hand,

$$A = \mathfrak{b}\mathfrak{a} \subseteq \mathfrak{a}^{-1}\mathfrak{a} \subseteq A,$$

so we must have $\mathfrak{a}^{-1}\mathfrak{a} = A$. □

EXAMPLE 11.10.16. Consider the maximal ideal (x,y) of $\mathbb{Q}[x,y]$. If $f \in \mathbb{Q}(x,y)^{\times}$ is such that $fx \in \mathbb{Q}[x,y]$ (resp., $fy \in \mathbb{Q}[x,y]$) then its denominator is a divisor of x (resp., y). Therefore $(x,y)^{-1} = \mathbb{Q}[x,y]$, and we have

$$(x,y)\cdot(x,y)^{-1} = (x,y) \neq \mathbb{Q}[x,y].$$

Thus, (x,y) is not invertible as a fractional ideal.

DEFINITION 11.10.17. A *principal fractional ideal* of A is an A-submodule (a) generated by a nonzero element a of the quotient field of A.

LEMMA 11.10.18. *Let $\mathfrak{a}$ be a fractional ideal of a PID. Then $\mathfrak{a}$ is principal.*

PROOF. There exists $d \in A$ such that $d\mathfrak{a} = (b)$ for some $b \in A$. Then $\frac{b}{d} \in \mathfrak{a}$ and given any $c \in \mathfrak{a}$, we have $dc = ba$ for some $a \in A$, so $c = a\frac{b}{d}$. That is, $\mathfrak{a} = (\frac{b}{d})$. □

LEMMA 11.10.19. *Let A be a domain, and let a be a nonzero element of its quotient field. Then (a) is invertible, and $(a)^{-1} = (a^{-1})$.*

PROOF. If $x \in (a)^{-1}$, then $xa = b$ for some $b \in A$, so $x = ba^{-1} \in (a^{-1})$. If $x \in (a^{-1})$, then $x = a^{-1}b$ for some $b \in A$. On other hand, any $z \in (a)$ has the form $z = ya$ for some $y \in A$, and we have $xz = a^{-1}bya = by \in A$, so $x \in (a)^{-1}$. We then have

$$(a)(a)^{-1} = (a)(a^{-1}) = (aa^{-1}) = A,$$

completing the proof. □

LEMMA 11.10.20. *Let A be a Dedekind domain, and let $\mathfrak{p}$ be a nonzero prime ideal of A. Then $\mathfrak{p}\mathfrak{p}^{-1} = A$.*

PROOF. Let $a \in \mathfrak{p}$ be nonzero. Noting Lemma 11.10.6a, we let $k \geq 1$ be minimal such that there exist nonzero prime ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_k$ of A with $\mathfrak{p}_1 \cdots \mathfrak{p}_k \subseteq (a)$. By Lemma 11.10.6b, we may without loss of generality suppose that $\mathfrak{p}_k = \mathfrak{p}$. By the minimality of k, we may choose $b \in \mathfrak{p}_1 \cdots \mathfrak{p}_{k-1}$ be such that $b \notin (a)$. Then $a^{-1}b \notin A$, but we have

$$a^{-1}b\mathfrak{p} \subseteq a^{-1}\mathfrak{p}_1 \cdots \mathfrak{p}_k \subseteq A,$$

which implies that $a^{-1}b \in \mathfrak{p}^{-1}$. Moreover, if $\mathfrak{p}^{-1}\mathfrak{p} = \mathfrak{p}$, then $a^{-1}b\mathfrak{p} \subseteq \mathfrak{p}$. Since $\mathfrak{p}$ is finitely generated, Proposition 11.3.4 tells us that $a^{-1}b$ is integral over A. But A is integrally closed, so we have a contradiction. That is, we must have $\mathfrak{p} \subsetneq \mathfrak{p}^{-1}\mathfrak{p} \subseteq A$, from which it follows that $\mathfrak{p}^{-1}\mathfrak{p} = A$ by maximality of $\mathfrak{p}$. □

THEOREM 11.10.21. *Let A be a Dedekind domain, and let $\mathfrak{a}$ be a fractional ideal of A. Then there exist $k \geq 0$ and distinct nonzero prime ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_k$ and $r_1, \dots, r_k \in \mathbb{Z} - \{0\}$ such that $\mathfrak{a} = \mathfrak{p}_1^{r_1} \cdots \mathfrak{p}_k^{r_k}$, and this decomposition is unique up to ordering. Moreover, $\mathfrak{a}$ is an ideal of A if and only if every r_i is positive.*

PROOF. First suppose that $\mathfrak{a}$ is a nonzero ideal of A. We work by induction on a nonnegative integer m such that there are nonzero prime ideals $\mathfrak{q}_1, \dots, \mathfrak{q}_m$ of $\mathfrak{a}$ (not necessarily distinct) with $\mathfrak{q}_1 \cdots \mathfrak{q}_m \subseteq \mathfrak{a}$, which exists by Lemma 11.10.6a. If $m = 0$, then $A \subseteq \mathfrak{a}$, so $\mathfrak{a} = A$. In general, for $m \geq 1$, we know that $\mathfrak{a}$ is proper, so there exists a nonzero prime ideal $\mathfrak{p}$ that contains $\mathfrak{a}$ and $\mathfrak{p} = \mathfrak{q}_i$ for some $i \leq m$. Without loss of generality, we take $i = m$. Then

$$\mathfrak{q}_1 \cdots \mathfrak{q}_{m-1} \subseteq \mathfrak{q}_1 \cdots \mathfrak{q}_m\mathfrak{p}^{-1} \subseteq \mathfrak{a}\mathfrak{p}^{-1} \subseteq A.$$

By induction, there exist nonzero prime ideals $\mathfrak{q}'_1, \dots, \mathfrak{q}'_\ell$ of A for some $\ell < m$ such that $\mathfrak{a}\mathfrak{p}^{-1} = \mathfrak{q}'_1 \cdots \mathfrak{q}'_\ell$. The desired factorization is given by multiplying by $\mathfrak{p}$, applying Lemma 11.10.20, and gathering together nondistinct primes.

In general, for a fractional ideal $\mathfrak{a}$, we let $d \in A$ be such that $d\mathfrak{a} \subseteq A$. We write $d\mathfrak{a} = \mathfrak{q}_1 \cdots \mathfrak{q}_m$ for some $m \geq 0$ and prime ideals $\mathfrak{q}_i$ for $1 \leq i \leq m$. We also write $(d) = \mathfrak{l}_1 \cdots \mathfrak{l}_n$ for some $n \geq 0$ and prime ideals $\mathfrak{l}_i$ for $1 \leq i \leq n$. By Lemma 11.10.20, we then have

$$\mathfrak{a} = (d)^{-1}(d\mathfrak{a}) = \mathfrak{l}_1^{-1} \cdots \mathfrak{l}_n^{-1}\mathfrak{q}_1 \cdots \mathfrak{q}_m.$$

If $\mathfrak{q}_i = \mathfrak{l}_j$ for some i and j, then we may use Lemma 11.10.20 to remove $\mathfrak{q}_i\mathfrak{l}_j^{-1}$ from the product. Hence we have the desired factorization.

Now suppose that

$$\mathfrak{a} = \mathfrak{p}_1^{r_1} \cdots \mathfrak{p}_k^{r_k}$$

for some $k \geq 0$, distinct primes $\mathfrak{p}_1, \dots, \mathfrak{p}_k$ and nonzero $r_1, \dots, r_k$. For each prime $\mathfrak{p}$ of A, consider the localization $A_\mathfrak{p}$, which is a Dedekind domain with unique nonzero prime ideal $\mathfrak{p}A_\mathfrak{p}$. Note that $\mathfrak{q}A_\mathfrak{p} = A_\mathfrak{p}$ if $\mathfrak{q}$ is a nonzero prime of A other than $\mathfrak{p}$. We therefore have

$$\mathfrak{a}A_\mathfrak{p} = \mathfrak{p}_1^{r_1} \dots, \mathfrak{p}_k^{r_k}A_\mathfrak{p} = \mathfrak{p}^r A_\mathfrak{p},$$

where $r = r_i$ if $\mathfrak{p} = \mathfrak{p}_i$ for some i, and $r = 0$ otherwise. Moreover, if $\mathfrak{p}^k A_{\mathfrak{p}} = \mathfrak{p}^l A_{\mathfrak{p}}$ for some integers $k \le l$, then $\mathfrak{p}^{l-k} A_{\mathfrak{p}} = A_{\mathfrak{p}}$, which since $\mathfrak{p}$ is nonzero, can only happen if $k = l$. Therefore, the primes $\mathfrak{p}_i$ and corresponding integers r_i are uniquely determined by $\mathfrak{a}$. □

We have the following immediate corollary of Theorem 11.10.21.

COROLLARY 11.10.22. *The set of fractional ideals $I(A)$ of a Dedekind domain A is a group under multiplication of fractional ideals with identity A, the inverse of $\mathfrak{a} \in I(A)$ being $\mathfrak{a}^{-1}$.*

DEFINITION 11.10.23. Let A be a Dedekind domain. The group $I(A)$ of fractional ideals of A is called the *ideal group* of A.

DEFINITION 11.10.24. Let A be a Dedekind domain. Then we let $P(A)$ denote the set of its principal fractional ideals. We refer to this as the *principal ideal group*.

COROLLARY 11.10.25. *Let A be a Dedekind domain. The group $P(A)$ is a subgroup of $I(A)$.*

DEFINITION 11.10.26. The *class group* (or *ideal class group*) of a Dedekind domain A is $\mathrm{Cl}(A) = I(A)/P(A)$, the quotient of the ideal group by the principal ideal group.

LEMMA 11.10.27. *A Dedekind domain A is a PID if and only if* $\mathrm{Cl}(A)$ *is trivial.*

PROOF. Every element of $I(A)$ has the form $\mathfrak{a}\mathfrak{b}^{-1}$ where $\mathfrak{a}$ and $\mathfrak{b}$ are nonzero ideals of A. If A is a PID, then both $\mathfrak{a}$ and $\mathfrak{b}$ are principal and, therefore, so is $\mathfrak{a}\mathfrak{b}^{-1}$. On the other hand, if $\mathfrak{a}$ is a nonzero ideal of A with $\mathfrak{a} = (a)$ for some $a \in K$, then clearly $a \in A$, so $\mathrm{Cl}(A)$ being trivial implies that A is a PID. □

NOTATION 11.10.28. Let K be a number field. We let I_K, P_K, and Cl_K denote the ideal group, principal ideal group, and class group of $\mathcal{O}_K$, respectively. We refer to these as the *ideal group* of K, the *principal ideal group* of K, and the *class group* of K, respectively.

EXAMPLE 11.10.29. Let $K = \mathbb{Q}(\sqrt{-5})$. Then $\mathcal{O}_K = \mathbb{Z}[\sqrt{-5}]$. The ideal $\mathfrak{a} = (2, 1+\sqrt{-5})$ is non-principal. To see this, note that $N_{K/\mathbb{Q}}(2) = 4$ and $N_{K/\mathbb{Q}}(1+\sqrt{-5}) = 6$, so any generator x of $\mathfrak{a}$ must satisfy $N_{K/\mathbb{Q}}(x) \in \{\pm 1, \pm 2\}$. But

$$N_{K/\mathbb{Q}}(a + b\sqrt{-5}) = a^2 + 5b^2$$

for $a, b \in \mathbb{Z}$, which forces $x = \pm 1$. This would mean that $\mathfrak{a} = \mathbb{Z}[\sqrt{-5}]$. To see that this cannot happen, define $\phi \colon \mathbb{Z}[\sqrt{-5}] \to \mathbb{Z}/6\mathbb{Z}$ by $\phi(a + b\sqrt{-5}) = a - b$ for $a, b \in \mathbb{Z}$. This is a ring homomorphism as

$$\begin{aligned}\phi((a+b\sqrt{-5})(c+d\sqrt{-5})) = \phi(ac-5bd+(ad+bc)\sqrt{-5}) &= ac-5bd-ad-bc \\ &= ac+bd-ad-bc = (a-b)(c-d).\end{aligned}$$

Moreover, $\phi(1+\sqrt{-5}) = 0$, so the kernel of ϕ contains (and is in fact equal to) $(1+\sqrt{-5})$. Therefore, ϕ induces a surjection (in fact, isomorphism),

$$\mathbb{Z}[\sqrt{-5}]/\mathfrak{a} \to \mathbb{Z}/6\mathbb{Z}/(2) \to \mathbb{Z}/2\mathbb{Z},$$

so $\mathfrak{a} \neq \mathbb{Z}[\sqrt{-5}]$, and x does not exist. Therefore, $\mathrm{Cl}_{\mathbb{Q}(\sqrt{-5})}$ is nontrivial.

We end with the following important theorem.

THEOREM 11.10.30. *A Dedekind domain is a UFD if and only if it is a PID.*

PROOF. We need only show that a Dedekind domain that is a UFD is a PID. Let A be such a Dedekind domain. By Theorem 11.10.21, it suffices to show that each nonzero prime ideal $\mathfrak{p}$ of A is principal. Since $\mathfrak{p}$ is prime and A is a UFD, any nonzero element of $\mathfrak{p}$ is divisible by an irreducible element in $\mathfrak{p}$. If π is such an element, then (π) is maximal and contained in $\mathfrak{p}$, so $\mathfrak{p} = (\pi)$. □

11.11. Discrete valuation rings

DEFINITION 11.11.1. A *discrete valuation ring*, or *DVR*, is a principal ideal domain that has exactly one nonzero prime ideal.

LEMMA 11.11.2. *The following are equivalent conditions on a principal ideal domain A.*

i. A is a DVR,

ii. A has a unique nonzero maximal ideal,

iii. A has a unique nonzero irreducible element up to associates.

PROOF. This is a simple consequence of the fact that in a PID, every nonzero prime ideal is maximal generated by any irreducible element it contains. □

DEFINITION 11.11.3. A *uniformizer* of a DVR is a generator of its maximal ideal.

Moreover, we have the following a priori weaker but in fact equivalent condition for a domain to be a DVR.

PROPOSITION 11.11.4. *A domain A is a DVR if and only if it is a local Dedekind domain that is not a field.*

PROOF. A DVR is a PID, hence a Dedekind domain, and it is local by definition. Conversely, suppose that A is noetherian, integrally closed, and has a unique nonzero prime ideal $\mathfrak{p}$. We must show that A is a PID. Since nonzero ideals factor uniquely as products of primes in A, every ideal of A has the form $\mathfrak{p}^n$ for some n. In particular, $\mathfrak{p} = (\pi)$ for any $\pi \in \mathfrak{p} - \mathfrak{p}^2$, and then $\mathfrak{p}^n = (\pi^n)$ for all n. Therefore, A is a PID and hence a DVR. □

THEOREM 11.11.5. *A noetherian domain A is a Dedekind domain if and only if its localization at every nonzero prime ideal is a DVR.*

PROOF. We have seen in Proposition 11.10.5 that $A_{\mathfrak{p}}$ is a Dedekind domain for all nonzero prime ideals $\mathfrak{p}$. By Proposition 11.11.4, each such localization is therefore a DVR.

Conversely, if A is a noetherian integral domain such that $A_{\mathfrak{p}}$ is a DVR for every nonzero prime ideal $\mathfrak{p}$, we consider the intersection $B = \bigcap_{\mathfrak{p}} A_{\mathfrak{p}}$ over all nonzero prime ideals $\mathfrak{p}$ of A, taken inside the quotient field K of A. Clearly, B contains A, and if $\frac{c}{d} \in B$ for some $c, d \in A$, then we set

$$\mathfrak{a} = \{a \in A \mid ac \in (d)\}.$$

By definition of B, we may write $\frac{c}{d} = \frac{r}{s}$ with $r \in A$ and $s \in A - \mathfrak{p}$, and we see that $sc = rd$, so $s \in \mathfrak{a}$. In other words, we have $\mathfrak{a} \not\subseteq \mathfrak{p}$ for all prime ideals $\mathfrak{p}$ of A, which forces $\mathfrak{a} = A$. This implies that $c \in (d)$, so $\frac{c}{d} \in A$.

Next, suppose that $\mathfrak{q}$ is a nonzero prime ideal of A, and let $\mathfrak{m}$ be a maximal ideal containing it. Then $\mathfrak{q}A_{\mathfrak{m}}$ is a nonzero prime ideal of $A_{\mathfrak{m}}$, which is a DVR, so $\mathfrak{q}A_{\mathfrak{m}} = \mathfrak{m}A_{\mathfrak{m}}$. Since $\mathfrak{q}$ and $\mathfrak{m}$ are prime ideals contained in $\mathfrak{m}$, we therefore have

$$\mathfrak{q} = A \cap \mathfrak{q}A_{\mathfrak{m}} = A \cap \mathfrak{m}A_{\mathfrak{m}} = \mathfrak{m}.$$

Thus, A has Krull dimension at most 1.

Finally, each $A_{\mathfrak{p}}$ is integrally closed in K by Corollary 11.3.22, and then the intersection A is as well, since any element of K that is integral over A is integral over each $A_{\mathfrak{p}}$, hence contained in each $A_{\mathfrak{p}}$. That is, A satisfies the conditions in the definition of a Dedekind domain. □

To make some sense of the name "discrete valuation ring", we define the notion of a discrete valuation. For this purpose, we adjoin an element ∞ to $\mathbb{Z}$ which is considered larger than any element of $\mathbb{Z}$, and we set $x + y = \infty$ if $x, y \in \mathbb{Z} \cup \{\infty\}$ and either x or y equals ∞.

DEFINITION 11.11.6. Let K be a field. A *discrete valuation* on K is a surjective map $v \colon K \to \mathbb{Z} \cup \{\infty\}$ such that

i. $v(a) = \infty$ if and only $a = 0$,

ii. $v(ab) = v(a) + v(b)$, and

iii. $v(a+b) \geq \min(v(a), v(b))$

for all $a, b \in K$.

DEFINITION 11.11.7. If v is a discrete valuation on a field K, then the quantity $v(a)$ for $a \in K$ is said to be the *valuation* of a with respect to v.

The following are standard examples of discrete valuations.

EXAMPLE 11.11.8. Let p be a prime number. Then the p-adic valuation v_p on $\mathbb{Q}$ is defined by $v_p(0) = \infty$ and $v_p(a) = r$ for $a \in \mathbb{Q}^{\times}$ if $a = p^r a'$ for some $r \in \mathbb{Z}$ and $a' \in \mathbb{Q}^{\times}$ such that p divides neither the numerator nor denominator of a' in reduced form.

EXAMPLE 11.11.9. Let F be a field, and consider the function field $F(t)$. The valuation at ∞ on $F(t)$ is defined by $v_{\infty}(\frac{g}{h}) = \deg h - \deg g$ for $g, h \in F[t]$ with $h \neq 0$, taking $\deg 0 = \infty$.

More generally, we have the following.

DEFINITION 11.11.10. Let A be a Dedekind domain with quotient field K, and let $\mathfrak{p}$ be a nonzero prime ideal of A. The $\mathfrak{p}$-*adic valuation* $v_{\mathfrak{p}}$ on K is defined on $a \in K^{\times}$ as the unique integer such that $(a) = \mathfrak{p}^{v_p(a)}\mathfrak{b}\mathfrak{c}^{-1}$ for some nonzero ideals $\mathfrak{b}$ and $\mathfrak{c}$ of A that are not divisible by $\mathfrak{p}$.

EXAMPLE 11.11.11. For the valuation at ∞ on $F(t)$, where F is a field, we may take $A = K[t^{-1}]$ and $\mathfrak{p} = (t^{-1})$. Then the valuation v_{∞} on $F(t)$ is the (t^{-1})-adic valuation. To see this, note that for nonzero $g,h \in F[t]$, one has

$$\frac{g(t)}{h(t)} = (t^{-1})^{\deg h - \deg g} \frac{G(t^{-1})}{H(t^{-1})},$$

where $G(t^{-1}) = t^{-\deg g} g(t)$ and $H(t^{-1}) = t^{-\deg h} h(t)$ are polynomials in t^{-1} which have nonzero constant term.

LEMMA 11.11.12. *Let A be a Dedekind domain with quotient field K, and let $\mathfrak{p}$ be a prime ideal of A. The $\mathfrak{p}$-adic valuation on K is a discrete valuation.*

PROOF. Let $a,b \in K$ be nonzero (without loss of generality). Write $(a) = \mathfrak{p}^r \mathfrak{a}$ and $(b) = \mathfrak{p}^s \mathfrak{b}$ for $r = v_{\mathfrak{p}}(a)$ and $s = v_{\mathfrak{p}}(b)$ and fractional ideals $\mathfrak{a}$ and $\mathfrak{b}$ of A. Note that $(ab) = \mathfrak{p}^{r+s}\mathfrak{a}\mathfrak{b}$, so $v_p(ab) = r+s$. We have

$$(a+b) = \mathfrak{p}^r\mathfrak{a} + \mathfrak{p}^s\mathfrak{b} = \mathfrak{p}^{\min(r,s)}(\mathfrak{p}^{r-\min(r,s)}\mathfrak{a} + \mathfrak{p}^{s-\min(r,s)}\mathfrak{b}),$$

so

$$v_{\mathfrak{p}}(a+b) = \min(r,s) + v_{\mathfrak{p}}(\mathfrak{p}^{r-\min(r,s)}\mathfrak{a} + \mathfrak{p}^{s-\min(r,s)}\mathfrak{b}) \geq \min(r,s).$$

□

LEMMA 11.11.13. *Let v be a discrete valuation on a field K. Then we have $v(-a) = v(a)$ for all $a \in K$.*

PROOF. Note that $2v(-1) = v(1) = 0$, so we have $v(-a) = v(-1) + v(a) = v(a)$. □

LEMMA 11.11.14. *Let v be a discrete valuation on a field K. Then we have*

$$v(a+b) = \min(v(a), v(b))$$

for all $a,b \in K$ with $v(a) \neq v(b)$.

PROOF. If $v(a) < v(b)$, then

$$v(a) = v((a+b)-b) \geq \min(v(a+b), v(b)) \geq \min(v(a), v(b)) = v(a),$$

so we have $v(a) = \min(v(a+b), v(b))$, which forces $v(a+b) = v(a)$. □

DEFINITION 11.11.15. Let K be a field, and let v be a discrete valuation on K. Then

$$\mathscr{O}_v = \{a \in K \mid v(a) \geq 0\}$$

is called the *valuation ring* of v.

LEMMA 11.11.16. *Let K be a field, and let v be a discrete valuation on K. Then $\mathscr{O}_v$ is a DVR with maximal ideal*

$$\mathfrak{m}_v = \{a \in K \mid v(a) \geq 1\}.$$

PROOF. That $\mathscr{O}_v$ is a ring follows from the fact that if $a,b \in \mathscr{O}_v$, then $v(ab) = v(a) + v(b) \geq 0$, $v(-a) = v(a) \geq 0$, and $v(a+b) \geq \min(v(a), v(b)) \geq 0$. For $a \in \mathscr{O}_v$ and $x,y \in \mathfrak{m}_v$, we have $v(x+y) \geq \min(v(x), v(y)) \geq 1$ and $v(ax) = v(a) + v(x) \geq 1$, so $\mathfrak{m}_v$ is an ideal. It is also the unique maximal ideal: given $a \in \mathscr{O}_v - \mathfrak{m}_v$, we have $v(a^{-1}) = v(a) + v(a^{-1}) = v(1) = 0$, so $a \in \mathscr{O}_v^{\times}$. Given an ideal $\mathfrak{a}$ of $\mathscr{O}_v$, let $a \in \mathfrak{a}$ be an element of minimal valuation n. Let $\pi \in \mathscr{O}_v$ with $v(\pi) = 1$, and write $a = \pi^n u$ for some $u \in \mathscr{O}_v^\times$. Then $v(u) = 0$, so $u \in \mathscr{O}_v^\times$. Therefore, $(\pi^n) \subseteq \mathfrak{a}$. On the other hand, since n is minimal, we have $\mathfrak{a} \subseteq (\pi^n)$, and therefore $\mathfrak{a}$ is a principal. By Lemma 11.11.2, we conclude that $\mathscr{O}_v$ is a DVR. □

EXAMPLE 11.11.17. In $\mathbb{Q}$, we have

$$\mathscr{O}_{v_p} = \mathbb{Z}_{(p)} = \left\{ \frac{a}{b} \mid a, b \in \mathbb{Z} \text{ such that } p \nmid b \right\}.$$

11.12. Ramification of primes

The integral closure B of a Dedekind domain A in a finite extension L of its quotient field K is also a Dedekind domain. If $\mathfrak{p}$ is a nonzero prime ideal of A, then we can consider the ideal $\mathfrak{p}B$ of B. This ideal may no longer be prime. Instead, it has a factorization

$$\mathfrak{p}B = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_g^{e_g} \tag{11.12.1}$$

for some distinct nonzero prime ideals $\mathfrak{P}_i$ of B and positive integers e_i, for $1 \le i \le g$ for some $g \ge 1$. We make the following definitions.

DEFINITION 11.12.1. Let B/A be an extension of commutative rings. We say that a prime ideal $\mathfrak{P}$ of B *lies over* (or *above*) a prime ideal $\mathfrak{p}$ of A if $\mathfrak{p} = \mathfrak{P} \cap A$. We then say that $\mathfrak{p}$ *lies under* (or *below*) $\mathfrak{P}$.

In (11.12.1), the prime ideals of B lying over $\mathfrak{p}$ are exactly the $\mathfrak{P}_i$ for $1 \le i \le g$.

DEFINITION 11.12.2. Let A be a Dedekind domain, and let B be the integral closure of A in a finite extension L of the quotient field K of A. Let $\mathfrak{p}$ be a nonzero prime ideal of A.

a. We say that $\mathfrak{p}$ *ramifies* (or *is ramified*) in L/K if $\mathfrak{p}B$ is divisible by the square of a prime ideal of B. Otherwise, it is said to be *unramified*.

b. We say that $\mathfrak{p}$ is *inert* in L/K if $\mathfrak{p}B$ is a prime ideal.

c. We say that $\mathfrak{p}$ is *split* in L/K if there exist two distinct prime ideals of B lying over $\mathfrak{p}$. Otherwise, $\mathfrak{p}$ is *non-split*.

It follows directly that $\mathfrak{p}$ is ramified in L/K if some e_i in (11.12.1) is at least 2. On the other hand, $\mathfrak{p}$ is inert in L/K if there is exactly one prime ideal of B lying over $\mathfrak{p}$ and its ramification index is 1, which is to say that $g = 1$ and $e_1 = 1$ in (11.12.1). Finally, $\mathfrak{p}$ is split in L/K if $g > 1$.

EXAMPLE 11.12.3. Let $A = \mathbb{Z}$ and $L = \mathbb{Q}(\sqrt{2})$. The integral closure of A in L is $B = \mathscr{O}_L = \mathbb{Z}[\sqrt{2}]$. The prime $\mathfrak{p} = (2)$ ramifies in $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$, since

$$2\mathbb{Z}[\sqrt{2}] = (\sqrt{2})^2.$$

Moreover, $\mathfrak{P} = (\sqrt{2})$ is a prime ideal of $\mathbb{Z}[\sqrt{2}]$, since $\mathbb{Z}[\sqrt{2}]/(\sqrt{2}) \cong \mathbb{Z}/2\mathbb{Z}$ via the map that takes $a + b\sqrt{2}$ to a mod 2. Therefore, $\mathfrak{p}$ is ramified and non-split.

Next, consider the prime ideal (3) of $\mathbb{Z}$. We have $\mathbb{Z}[\sqrt{2}]/(3) \cong \mathbb{F}_3[\sqrt{2}] \cong \mathbb{F}_9$, so (3) is inert in $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$. On the other hand, the prime factorization of $7\mathbb{Z}[\sqrt{2}]$ is exactly

$$7\mathbb{Z}[\sqrt{2}] = (3 + \sqrt{2})(3 - \sqrt{2}),$$

since $\mathbb{Z}[\sqrt{2}]/(3 \pm \sqrt{2})$ is isomorphic to $\mathbb{Z}/7\mathbb{Z}$ via the map that takes $a + b\sqrt{2}$ to $a \mp 3b$. That is, (7) splits in $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$.

DEFINITION 11.12.4. Let A be a Dedekind domain, and let $\mathfrak{p}$ be a nonzero prime ideal of A. The *residue field* of $\mathfrak{p}$ is $A/\mathfrak{p}$.

REMARK 11.12.5. Let A be a Dedekind domain, and let B be the integral closure of A in a finite extension L the quotient field K of A. Let $\mathfrak{p}$ be a nonzero prime ideal of A, and let $\mathfrak{P}$ be a prime ideal of L lying over K. Then $B/\mathfrak{P}$ is a field extension of $A/\mathfrak{p}$ via the natural map induced on quotients by the inclusion $A \hookrightarrow B$.

DEFINITION 11.12.6. Let A be a Dedekind domain, and let B be the integral closure of A in a finite extension L of the quotient field K of A. Let $\mathfrak{p}$ be a nonzero prime ideal of A, and let $\mathfrak{P}$ be a prime ideal of B lying over $\mathfrak{p}$.

a. The *ramification index* $e_{\mathfrak{P}/\mathfrak{p}}$ of $\mathfrak{P}$ over $\mathfrak{p}$ is the largest $e \geq 1$ such that $\mathfrak{P}^e$ divides $\mathfrak{p}B$.

b. The *residue degree* $f_{\mathfrak{P}/\mathfrak{p}}$ of a prime ideal of $\mathfrak{P}$ lying over $\mathfrak{p}$ is $[B/\mathfrak{P} : A/\mathfrak{p}]$.

REMARK 11.12.7. It follows quickly from the definitions that ramification indices and residue degrees are multiplicative in extensions. That is, if $A \subseteq B \subseteq C$ are Dedekind domains with the quotient field of C a finite extension of that of A and $\mathfrak{P}$ is a prime ideal of C lying over P of B and $\mathfrak{p}$ of A, then

$$e_{\mathfrak{P}/\mathfrak{p}} = e_{\mathfrak{P}/P}e_{P/\mathfrak{p}} \quad \text{and} \quad f_{\mathfrak{P}/\mathfrak{p}} = f_{\mathfrak{P}/P}f_{P/\mathfrak{p}}.$$

EXAMPLE 11.12.8. In Example 11.12.3, the residue degree of $(\sqrt{2})$ over $2\mathbb{Z}$ is 1, the residue degree of $3\mathbb{Z}[\sqrt{2}]$ over $3\mathbb{Z}$ is 2, and the residue degrees of $(3 \pm \sqrt{2})$ over $7\mathbb{Z}$ are each 1. The ramification indices are 2, 1, and 1, repsectively.

We shall require the following lemmas.

LEMMA 11.12.9. *Let $\mathfrak{p}$ be a nonzero prime ideal in a Dedekind domain A. For each $i \geq 0$, the $A/\mathfrak{p}$-vector space $\mathfrak{p}^i/\mathfrak{p}^{i+1}$ is one-dimensional.*

PROOF. Let $x \in \mathfrak{p}^i - \mathfrak{p}^{i+1}$ for some $i \geq 0$. (Such an element exists by unique factorization of ideals.) We need only show that the image of x spans $\mathfrak{p}^i/\mathfrak{p}^{i+1}$. For this, note that $(x) = \mathfrak{p}^i\mathfrak{a}$ for some nonzero ideal of A not divisible by $\mathfrak{p}$. Then

$$(x) + \mathfrak{p}^{i+1} = \mathfrak{p}^i(\mathfrak{a} + \mathfrak{p}) = \mathfrak{p}^i,$$

the last step by the Chinese remainder theorem. □

LEMMA 11.12.10. *Let A be a Dedekind domain and P be a set of nonzero prime ideals of A. Let S a multiplicatively closed subset of A such that $S \cap \mathfrak{p} = \varnothing$ for all $\mathfrak{p} \in P$. Let $\mathfrak{a}$ be a nonzero ideal of A that is divisible only by prime ideals in P. Then the natural map*

$$A/\mathfrak{a} \to S^{-1}A/S^{-1}\mathfrak{a}$$

is an isomorphism.

PROOF. Suppose that $b \in S^{-1}\mathfrak{a} \cap A$, and write $b = \frac{a}{s}$ for some $a \in \mathfrak{a}$ and $s \in S$. Then $a = bs$, and since $\mathfrak{a}$ divides (a) while (s) is relatively prime to $\mathfrak{a}$, we must have that $\mathfrak{a}$ divides (b). In other words, $b \in \mathfrak{a}$, and therefore the map is injective. Given $c \in A$ and $t \in S$, the ideals (t) and $\mathfrak{a}$ have no common prime factor, so in that A is a Dedekind domain, satisfy $(t) + \mathfrak{a} = A$. Thus, there exists $u \in A$ such that $ut - 1 \in \mathfrak{a}$. Then $cu + \mathfrak{a}$ maps to $\frac{c}{t} + S^{-1}\mathfrak{a}$, so the map is surjective. □

The ramification indices and residue degrees of the primes over $\mathfrak{p}$ satisfy the following degree formula.

THEOREM 11.12.11. *Let A be a Dedekind domain, and let B be the integral closure of A in a finite separable extension L the quotient field K of A. Let $\mathfrak{p}$ be a nonzero prime ideal of A, and write*

$$\mathfrak{p}B = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_g^{e_g}$$

for some distinct nonzero prime ideals $\mathfrak{P}_i$ of B and positive integers e_i, for $1 \le i \le g$ and some $g \ge 1$. For each i, let $f_i = f_{\mathfrak{P}_i/\mathfrak{p}}$. Then

$$\sum_{i=1}^{g} e_i f_i = [L : K].$$

PROOF. We prove that $\dim_{A/\mathfrak{p}} B/\mathfrak{p}B$ equals both quantities in the desired equality. By the Chinese remainder theorem, we have a canonical isomorphism

$$B/\mathfrak{p}B \cong \prod_{i=1}^{g} B/\mathfrak{P}_i^{e_i},$$

of $A/\mathfrak{p}$-vector spaces, so

$$\dim_{A/\mathfrak{p}} B/\mathfrak{p}B = \sum_{i=1}^{g} \dim_{A/\mathfrak{p}} B/\mathfrak{P}_i^{e_i} = \sum_{i=1}^{g} \sum_{j=0}^{e_i-1} \dim_{A/\mathfrak{p}} \mathfrak{P}_i^{j}/\mathfrak{P}_i^{j+1}.$$

By Lemma 11.12.9, each $\mathfrak{P}_i^{j}/\mathfrak{P}_i^{j+1}$ is a 1-dimensional $B/\mathfrak{P}_i$-vector space, and we therefore have

$$\dim_{A/\mathfrak{p}} B/\mathfrak{p}B = \sum_{i=1}^{g} e_i \dim_{A/\mathfrak{p}} B/\mathfrak{P}_i = \sum_{i=1}^{g} e_i f_i.$$

Let S denote the complement of $\mathfrak{p}$ in A. Then $S^{-1}A = A_{\mathfrak{p}}$ and $S^{-1}B$ are Dedekind domains, and $A_{\mathfrak{p}}$ is a DVR, hence a PID. Moreover, $S^{-1}B$ is the integral closure of $A_{\mathfrak{p}}$ in L, being both integrally closed and contained in said integral closure. Thus, Corollary 11.3.39 tells us that $S^{-1}B$ is free of rank $[L : K]$ over $A_{\mathfrak{p}}$. In particular, $S^{-1}B/\mathfrak{p}S^{-1}B$ is an $[L : K]$-dimensional $A_{\mathfrak{p}}/\mathfrak{p}A_{\mathfrak{p}}$-vector space. On the other hand, note that

$$S \cap \mathfrak{P}_i = S \cap A \cap \mathfrak{P}_i = S \cap \mathfrak{p} = \varnothing$$

for each $1 \le i \le g$. Therefore, Lemma 11.12.10 tells us that

$$S^{-1}B/\mathfrak{p}S^{-1}B \cong B/\mathfrak{p}B$$

and $A_{\mathfrak{p}}/\mathfrak{p}A_{\mathfrak{p}} \cong A/\mathfrak{p}$. We thus have that $\dim_{A/\mathfrak{p}} B/\mathfrak{p}B = [L : K]$, as required. □

In other words, Theorem 11.12.11 tells us that the sum over all primes lying over $\mathfrak{p}$ of the products of their ramification indices with their residue degrees equals the degree of the field extension L/K.

CHAPTER 12

Homological algebra

We work in this chapter largely in an abelian category. At times, proofs of statements that hold true in arbitrary abelian categories will be given only in categories of modules over a ring. This choice, which simplifies the exposition, will be given a more rigorous justification in the course of the chapter.

12.1. Exact sequences

Though we've managed to suppress them to this point, exact sequences are ubiquitous in algebra. Let's give the definitions.

DEFINITION 12.1.1. Let I be the set of integers in an interval in $\mathbb{R}$. A diagram $A.$ in a category $\mathscr{C}$ of the form

$$\cdots \to A_{i+1} \xrightarrow{d_{i+1}^A} A_i \xrightarrow{d_i^A} A_{i-1} \to \cdots$$

is a *sequence*, where the A_i are defined for $i \in I$ and the morphisms $d_i^A \colon A_i \to A_{i-1}$ are defined for $i \in I$ with $i-1 \in I$. We will refer to I as the defining interval of the sequence $A.$.

NOTATION 12.1.2. In an abelian category, if A is a subobject of an object B, we write $A \subseteq B$ to denote this and B/A for the cokernel of the inclusion morphism $A \to B$. For $f \colon B \to C$, we will let $f(A)$ denote the image of the composite of the inclusion with f.

We are particularly interested in exact sequences.

DEFINITION 12.1.3. We say that a diagram

$$A \xrightarrow{f} B \xrightarrow{g} C$$

in an abelian category $\mathscr{C}$ is *exact* if $g \circ f = 0$ and the induced monomorphism $\operatorname{im} f \to \ker g$ is an isomorphism.

DEFINITION 12.1.4. A sequence $A. = (A_i, d_i^A)$ with defining interval I in an abelian category $\mathscr{C}$ is *exact*, or an *exact sequence*, if the subdiagram

$$A_{i+1} \xrightarrow{d_i} A_i \xrightarrow{d_{i-1}} A_{i-1}$$

is exact for each $i \in I \cap (I+1) \cap (I-1)$.

That is, $A.$ is exact if $d_i \circ d_{i+1} = 0$ and the canonical morphism $\operatorname{im} d_{i+1} \to \ker d_i$ is an isomorphism for all i for which d_{i+1} and d_i are defined: we write this more simply as the identification $\operatorname{im} d_{i+1} = \ker d_i$ for subobjects of A_i.

REMARK 12.1.5. One can make the same definitions of exact sequences in the category of groups, or more generally in any "semi-abelian" category, and much of the discussion that follows remains the same.

REMARK 12.1.6. If the interval I of definition of $A.$ has a left (resp., right) endpoint N such that $A_N = 0$, then one can extend $A.$ to the left (resp., right) by taking $A_i = 0$ for all $i < N$ (resp., $i > N$). In fact, we could do this for any sequence for which I has an endpoint (without the condition $A_N = 0$), but we do not as the operation does not preserve exactness (nor do the to-be-defined morphisms between two sequences defined on different intervals extend to morphisms under this operation).

DEFINITION 12.1.7. Let $\mathscr{C}$ be an abelian category.

a. A *short exact sequence* in $\mathscr{C}$ is an exact sequence in $\mathscr{C}$ of the form

$$0 \to A \to B \to C \to 0.$$

b. A *left short exact sequence* in $\mathscr{C}$ is an exact sequence in $\mathscr{C}$ of the form

$$0 \to A \to B \to C.$$

c. A *right short exact sequence* in $\mathscr{C}$ is an exact sequence in $\mathscr{C}$ of the form

$$A \to B \to C \to 0.$$

REMARK 12.1.8. To say that

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

is exact is to say that f is a monomorphism, $\operatorname{im} f = \ker g$, and g is an epimorphism.

EXAMPLE 12.1.9. Multiplication-by-n provides a short exact sequence of abelian groups

$$0 \to \mathbb{Z} \xrightarrow{n} \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z} \to 0.$$

REMARK 12.1.10. If $f\colon A \to B$ is any morphism in an abelian category, then we have an exact sequence

$$0 \to \ker f \to A \xrightarrow{f} B \to \operatorname{coker} f \to 0.$$

Note that this provides two short exact sequences

$$0 \to \ker f \to A \to \operatorname{im} f \to 0 \quad \text{and} \quad 0 \to \operatorname{im} f \to B \to \operatorname{coker} f \to 0,$$

since $\operatorname{coim} f \to \operatorname{im} f$ is an isomorphism. We can "splice these back together" to get the 4-term sequence by taking the composite $A \to \operatorname{im} f \to B$, which is f.

DEFINITION 12.1.11. A *long exact sequence* in an abelian category $\mathscr{C}$ is an exact sequence $A. = (A_i, d_i^A)_{i \in \mathbb{Z}}$.

Frequently, a long exact sequence is expressed in the form

$$\cdots \to A_2 \to A_1 \to A_0 \to 0,$$

and we can extend it to all integers by setting $A_i = 0$ for all $i \leq -1$.

EXAMPLE 12.1.12. The sequence

$$\cdots \to \mathbb{Q}^2 \xrightarrow{f} \mathbb{Q}^2 \xrightarrow{f} \mathbb{Q}^2 \xrightarrow{f} \mathbb{Q}^2 \to \cdots$$

with $f\colon \mathbb{Q}^2 \to \mathbb{Q}^2$ defined by $f(a,b) = (0,a)$ for $a,b \in \mathbb{Q}$ is a long exact sequence of $\mathbb{Q}$-vector spaces. The sequence

$$0 \to \mathbb{Q} \xrightarrow{\iota_2} \mathbb{Q}^2 \xrightarrow{f} \mathbb{Q}^2 \xrightarrow{f} \mathbb{Q}^2 \to \cdots$$

of $\mathbb{Q}$-vector spaces with $\iota_2(b) = (0,b)$ is also a long exact sequence.

We next study maps between sequences. Let us make a formal definition.

DEFINITION 12.1.13. Let $A_\cdot = (A_i, d_i^A)$ and $B_\cdot = (B_i, d_i^B)$ be sequences in a category $\mathscr{C}$ with defining intervals I and J, respectively. A *morphism of sequences* $f_\cdot\colon A_\cdot \to B_\cdot$ in a category is a collection $(f_i)_{i \in I \cap J}$ of morphisms $f_i\colon A_i \to B_i$ in $\mathscr{C}$ such that $d_i^B \circ f_i = f_{i-1} \circ d_i^A$ for all $i \in I \cap J \cap (I+1) \cap (J+1)$.

REMARK 12.1.14. We can view the condition for a sequence of maps $f_i\colon A_i \to B_i$ between the terms of sequences $A_\cdot$ and $B_\cdot$ defined at all $i \in \mathbb{Z}$ to be a map of sequences as saying that the diagram

$$\begin{tikzcd}
\cdots \arrow[r] & A_{i+1} \arrow[r, "d_{i+1}^A"] \arrow[d, "f_{i+1}"] & A_i \arrow[r, "d_i^A"] \arrow[d, "f_i"] & A_{i-1} \arrow[r] \arrow[d, "f_{i-1}"] & \cdots \\
\cdots \arrow[r] & A_{i+1} \arrow[r, "d_{i+1}^B"] & A_i \arrow[r, "d_i^B"] & A_{i-1} \arrow[r] & \cdots
\end{tikzcd}$$

commutes.

12.2. The snake and five lemmas

The following result on maps between short exact sequences is the key to much of homological algebra.

THEOREM 12.2.1 (Snake lemma). *Let $\mathscr{C}$ be an abelian category, and let*

$$\begin{tikzcd}
0 \arrow[r] & A \arrow[r, "f"] \arrow[d, "\alpha"] & B \arrow[r, "g"] \arrow[d, "\beta"] & C \arrow[r] \arrow[d, "\gamma"] & 0 \\
0 \arrow[r] & A' \arrow[r, "f'"] & B' \arrow[r, "g'"] & C' \arrow[r] & 0.
\end{tikzcd}$$

be a commutative diagram in $\mathscr{C}$ with exact rows. Then there is an exact sequence

$$0 \to \ker f \to \ker \alpha \to \ker \beta \to \ker \gamma \xrightarrow{\delta} \operatorname{coker} \alpha \to \operatorname{coker} \beta \to \operatorname{coker} \gamma \to \operatorname{coker} g' \to 0$$

such that the resulting diagram

$$\begin{tikzcd}
 & & \ker\alpha \arrow[r] \arrow[d] & \ker\beta \arrow[r] \arrow[d] & \ker\gamma \arrow[d] \arrow[dddll, "\delta"] & & \\
0 \arrow[r] & \ker f \arrow[r] \arrow[ur] & A \arrow[r, "f"] \arrow[d, "\alpha"] & B \arrow[r, "g"] \arrow[d, "\beta"] & C \arrow[r] \arrow[d, "\gamma"] & 0 & \\
 & 0 \arrow[r] & A' \arrow[r, "f'"] \arrow[d] & B' \arrow[r, "g'"] \arrow[d] & C' \arrow[r] \arrow[d] & \operatorname{coker} g' \arrow[r] & 0 \\
 & & \operatorname{coker}\alpha \arrow[r] & \operatorname{coker}\beta \arrow[r] & \operatorname{coker}\gamma \arrow[ur] & &
\end{tikzcd}$$

with the natural inclusion and quotient maps is commutative.

PROOF. We work in the category of modules over a ring R. We define δ as follows. For $c \in \ker\gamma$, find $b \in B$ with $g(b) = c$. Then $g' \circ \beta(b) = \gamma(c) = 0$, so $\beta(b) = f'(a')$ for some $a' \in A'$. Let $\delta(c)$ denote the image $\bar{a}'$ of a' in $\operatorname{coker}\alpha$. To see that this is well-defined, note that if $b_2 \in B$ also satisfies $g(b_2) = c$, then $g(b - b_2) = 0$, so $b_2 - b = f(a)$ for some $a \in A$. We then have

$$\beta(b_2) = \beta(b) + \beta \circ f(a) = \beta(b) + f' \circ \alpha(a),$$

so $b_2 = f'(a' + \alpha(a))$. But $a' + \alpha(a)$ has image $\bar{a}'$ in $\operatorname{coker}\alpha$, so δ is well-defined. That δ is an R-module homomorphism follows easily from the construction.

We now check that the other maps are well-defined. Since

$$\beta \circ f(\ker\alpha) = f' \circ \alpha(\ker\alpha) = 0,$$

we have $f(\ker\alpha) \subseteq \ker\beta$. Similarly, $g(\ker\beta) \subset \ker\gamma$. Also, if $\bar{a}' \in \operatorname{coker}\alpha$, then we may lift it to $a' \in A'$, map to $b' \in B'$, and then project to $\bar{b}' \in \operatorname{coker}\beta$. This is well-defined as any other choice of a' differs by some $a \in A$, which causes $\bar{b}'$ to change by the image of $\beta(f(a))$, which is zero. Thus f' induces a well-defined homomorphism

$$\bar{f}' \colon \operatorname{coker}\alpha \to \operatorname{coker}\beta.$$

Similarly, we have a well-defined surjection

$$\bar{g}' \colon \operatorname{coker}\beta \to \operatorname{coker}\gamma.$$

We next check that our sequence is a complex. Note that $g \circ f = 0$, so the same is true on $\ker\alpha$, and $g' \circ f' = 0$, so $\bar{g}' \circ \bar{f}' = 0$ as well. Let $b \in \ker\beta$. Then $\delta(g(b))$ is given by considering $\beta(b) = 0$, lifting it to some $a' \in A'$, which we may take to be 0, and projecting to $\operatorname{coker}\alpha$. Hence $\delta(g(\ker\beta)) = 0$. On the other hand, if $c \in \ker\gamma$, then $\bar{f}'(\delta(c))$ is given by definition by projecting $\beta(b)$ to $\operatorname{coker}\beta$, where $g(b) = c$, hence is zero. Hence, the image of one map is contained in the kernel of the next at each term of the six term sequence.

Finally, we check exactness at each term. If $a \in \ker\alpha$, then $f(a) = 0$ implies $a \in \ker f$. Inclusion then provides a map $\ker f \to \ker\alpha$ that is by definition injective. If $b \in \ker\beta \cap \ker g$, then there exists $a \in A$ with $f(a) = b$. Since $f' \circ \alpha(a) = \beta \circ f(a) = 0$ and f' is injective, we have $\alpha(a) = 0$, or $a \in \ker \alpha$. Hence

$$f(\ker \alpha) = \ker(\ker \beta \to \ker \gamma),$$

and we have exactness at $\ker \beta$.

If $c \in \ker \delta$, then whenever $g(b) = c$ and $f'(a') = \beta(b)$, we have $\bar{a}' = 0$, letting $\bar{a}'$ denote the image of $a' \in \operatorname{coker} \alpha$. We then have $a' = f(a)$ for some $a \in A$, so $b_2 = b - f(a)$ still satisfies $f(b_2) = c$, but $\beta(b_2) = 0$. So $b_2 \in \ker \beta$, and we have exactness at $\ker \gamma$.

If $\bar{a}' \in \operatorname{coker} \alpha$ is the image of $a' \in A'$ and $\bar{f}'(\bar{a}') = 0$, then there exists $b \in B$ with $\beta(b) = f'(a')$. Now

$$\gamma(g(b)) = g'(\beta(b)) = g'(f'(a')) = 0,$$

so $g(b) \in \ker \gamma$, and $\delta(g(b)) = \bar{a}'$. Hence, we have exactness at $\operatorname{coker} \alpha$.

If $\bar{b}' \in \operatorname{coker} \beta$ is the image of $b' \in B'$ and $\bar{g}'(\bar{b}') = 0$, then there exists $c \in C$ with $g(b') = \gamma(c)$. Now $c = g(b)$ for some $b \in B$. And $b'_2 = b' - b$ has image $\bar{b}'$ in $\operatorname{coker} \beta$. On the other hand, $f'(b'_2) = 0$, so $b'_2 = f'(a')$ for some $a' \in A'$. If $\bar{a}' \in \operatorname{coker} \alpha$ is the image of a', then $\bar{f}'(\bar{a}') = \bar{b}'$ as the image of b'_2 in $\operatorname{coker} \beta$. Thus, we have exactness at $\operatorname{coker} \beta$. Finally, if $\bar{c}' \in \operatorname{coker} \gamma$ is the image of $c' \in C'$ with trivial image in $\operatorname{coker} g'$, then $c' = g'(b')$ for some $b' \in B'$, then the image $\bar{b}' \in \operatorname{coker} \beta$ of b' satisfies $\bar{g}'(\bar{b}') = \bar{c}'$. □

Next, we state another useful result on maps between exact sequences, known as the five lemma.

THEOREM 12.2.2 (Five lemma). *Let*

$$\begin{tikzcd}
A \arrow[r, "e"] \arrow[d, "\alpha"] & B \arrow[r, "f"] \arrow[d, "\beta"] & C \arrow[r, "g"] \arrow[d, "\gamma"] & D \arrow[r, "h"] \arrow[d, "\delta"] & E \arrow[d, "\varepsilon"] \\
A' \arrow[r, "e'"] & B' \arrow[r, "f'"] & C' \arrow[r, "g'"] & D' \arrow[r, "h'"] & E'.
\end{tikzcd}$$

be a commutative diagram with exact rows in an abelian category $\mathscr{C}$.

a. If β and δ are epimorphisms and ε is a monomorphism, then γ is an epimorphism.

b. If β and δ are monomorphisms and α is an epimorphism, then γ is a monomorphism.

c. If β and δ are isomorphisms, α is an epimorphism, and ε is a monomorphism, then γ is an isomorphism.

PROOF. We work in the category of modules over a ring R. It is immediate that parts a and b imply part c (the actual five lemma). We prove part a and note that it, if proven in an arbitrary abelian category, implies b in the opposite category, which is also abelian. Suppose that β and δ are surjective and ε is injective. Let $c' \in C'$, and note that $g'(c') = \delta(d)$ for some $d \in D$ by surjectivity of δ. Also,

$$\varepsilon(h(d)) = h'(\delta(d)) = h'(g'(c')) = 0,$$

so $h(d) = 0$ by injectivity of ε. By exactness of the top row at D, we then have $c \in C$ such that $g(c) = d$. Now,

$$g'(\gamma(c) - c') = \delta(g(c)) - g'(c') = \delta(d) - \delta(d) = 0,$$

so by exactness of the bottom row at C, there exists $b' \in B'$ such that $f'(b') = c' - \gamma(c)$. As β is surjective, there also exists $b \in B$ such that $\beta(b) = b'$. Set $x = c + f(b) \in C$. Then

$$\gamma(x) - \gamma(c) = \gamma(f(b)) = f'(\beta(b)) = f'(b') = c' - \gamma(c)$$

so $\gamma(c) = c'$. Thus, γ is surjective and part a is proven. □

12.3. Homology and cohomology

DEFINITION 12.3.1. Let $\mathscr{C}$ be an abelian category.

a. A *chain complex*, or more simply *complex*, in $\mathscr{C}$ is a sequence $A. = (A_i, d_i^A)_{i\in\mathbb{Z}}$ in $\mathscr{C}$ such that $d_i^A \circ d_{i+1}^A = 0$ for all $i \in \mathbb{Z}$.

b. For a chain complex $A.$ and $i \in \mathbb{Z}$, the morphism $d_i^A \colon A_i \to A_{i-1}$ is called the ith *differential* in the complex $A.$.

NOTATION 12.3.2. Unless otherwise specified, the ith object in a chain complex $A.$ will be denoted A_i and the ith differential by $d_i \colon A_i \to A_{i-1}$. If we have multiple complexes, we will use d_i^A to specify the differential on $A.$.

REMARK 12.3.3. Unlike with sequences in general (or exact sequences in particular), if the terms and morphisms of complex are specified only for some interval of integers, then we complete it to a complex by declaring all remaining objects and morphisms to be zero.

DEFINITION 12.3.4. A *morphism of complexes* in an abelian category is a morphism of sequences between complexes.

DEFINITION 12.3.5. The *category of chain complexes* $\mathbf{Ch}(\mathscr{C})$ for an abelian category $\mathscr{C}$ is the category with objects the complexes $(A_i, d_i^A)_{i\in\mathbb{Z}}$ in $\mathscr{C}$ and morphisms the morphisms of complexes in $\mathscr{C}$.

REMARK 12.3.6. A sequence of complexes is exact in $\mathbf{Ch}(\mathscr{C})$ if and only if it the resulting sequence of objects in each fixed degree is exact in $\mathscr{C}$. For instance, a sequence of complexes

$$0 \to A. \xrightarrow{f.} B. \xrightarrow{g.} C. \to 0$$

in $\mathbf{Ch}(\mathscr{C})$ is short exact if and only if each

$$0 \to A_i \xrightarrow{f_i} B_i \xrightarrow{g_i} C_i \to 0$$

is a short exact sequence.

Note that the category of chain complexes in $\mathscr{C}$ is a fully faithful subcategory of the category of sequences with defining interval $\mathbb{Z}$.

DEFINITION 12.3.7. A complex is often said to be *acyclic* if it is an exact sequence.

The reader can easily check the following.

PROPOSITION 12.3.8. *Let $\mathscr{C}$ be an abelian category. Then the category* $\mathbf{Ch}(\mathscr{C})$ *is an abelian category as well.*

DEFINITION 12.3.9. The *ith homology* of a complex $A.$ in an abelian category is the object

$$H_i(A) = \frac{\ker d_i^A}{\operatorname{im} d_{i+1}^A}.$$

REMARK 12.3.10. A complex $A.$ is exact if and only if $H_i(A) = 0$ for all $i \in \mathbb{Z}$.

EXAMPLE 12.3.11. The complex

$$\cdots \to \mathbb{Z}/8\mathbb{Z} \xrightarrow{4} \mathbb{Z}/8\mathbb{Z} \xrightarrow{4} \mathbb{Z}/8\mathbb{Z} \to \cdots$$

of abelian groups has ith homology group $2\mathbb{Z}/4\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z}$ for every i.

LEMMA 12.3.12. *Any morphism $f.\colon A. \to B.$ of complexes in $\mathscr{C}$ induces natural morphisms*

$$f_i^*\colon H_i(A) \to H_i(B)$$

for each $i \in \mathbb{Z}$. More specifically, these satisfy

$$\bar{\iota}_i^B \circ f_i^* \circ \bar{\pi}_i^A = \pi_i^B \circ f_i \circ \iota_i^A,$$

where $\pi_i^B\colon B_i \to \operatorname{coker} d_{i+1}^B$ and $\bar{\pi}_i^A\colon \ker d_i^A \to H_i(A)$ are the canonical epimorphisms, and where $\iota_i^A\colon \ker d_i^A \to A_i$ and $\bar{\iota}_i^B\colon H_i(B) \to \operatorname{coker} d_{i+1}^B$ are the canonical monomorphisms.

PROOF. We prove this in the case that $\mathscr{C}$ is a category of R-modules. If $a \in \ker d_i^A$, then $d_i^B(f_i(a)) = f_{i-1}(d_i^A(a)) = 0$, so $f_i(a) \in \ker d_i^B$. If $a = d_{i+1}^A(a')$, then $f_i(a) = d_{i+1}^B(f_{i+1}(a')) \in \operatorname{im} d_{i+1}^B$. Thus, the composite map

$$\ker d_i^A \xrightarrow{f_i} \ker d_i^B \twoheadrightarrow H_i(B)$$

factors through $H_i(A)$, inducing the stated map f_i^*. □

THEOREM 12.3.13. *Let $\mathscr{C}$ be an abelian category, and let*

$$0 \to A. \xrightarrow{f.} B. \xrightarrow{g.} C. \to 0$$

be a short exact sequence in $\mathbf{Ch}(\mathscr{C})$. There there are morphisms $\delta_i\colon H_i(C) \to H_{i-1}(A)$ for all $i \in \mathbb{Z}$, natural in the short exact sequence, that fit into a long exact sequence

$$\cdots \to H_i(A) \xrightarrow{f_i^*} H_i(B) \xrightarrow{g_i^*} H_i(C) \xrightarrow{\delta_i} H_{i-1}(A) \to \cdots.$$

PROOF. First, the snake lemma applied to the diagram

$$\begin{tikzcd}
0 \arrow[r] & A_i \arrow[r, "f_i"] \arrow[d, "d_i^A"] & B_i \arrow[r, "g_i"] \arrow[d, "d_i^B"] & C_i \arrow[r] \arrow[d, "d_i^C"] & 0 \\
0 \arrow[r] & A_{i-1} \arrow[r, "f_{i-1}"] & B_{i-1} \arrow[r, "g_{i-1}"] & C_{i-1} \arrow[r] & 0
\end{tikzcd}$$

provides exact sequences

$$0 \to \ker d_i^A \to \ker d_i^B \to \ker d_i^C \quad \text{and} \quad \operatorname{coker} d_i^A \to \operatorname{coker} d_i^B \to \operatorname{coker} d_i^C \to 0.$$

Then, we wish to apply the snake lemma to the diagram

$$\begin{tikzcd}
 & \operatorname{coker} d^A_{i+1} \arrow[r, "\bar{f}_i"] \arrow[d, "\bar{d}^A_i"] & \operatorname{coker} d^B_{i+1} \arrow[r, "\bar{g}_i"] \arrow[d, "\bar{d}^B_i"] & \operatorname{coker} d^C_i \arrow[r] \arrow[d, "\bar{d}^C_i"] & 0 \\
0 \arrow[r] & \ker d^A_{i-1} \arrow[r, "f_{i-1}"] & \ker d^B_{i-1} \arrow[r, "g_{i-1}"] & \ker d^C_{i-1} &
\end{tikzcd}$$

with exact rows (where the "bars" denote morphisms induced on quotients). By the snake lemma, we have an exact sequence

$$\ker \bar{d}^A_i \to \ker \bar{d}^B_i \to \ker \bar{d}^C_i \xrightarrow{\delta_i} \operatorname{coker} \bar{d}^A_i \to \operatorname{coker} \bar{d}^B_i \to \operatorname{coker} \bar{d}^C_i .$$

Note that for $X \in \{A, B, C\}$, we have

$$\ker \bar{d}^X_i = \ker(X_i / \operatorname{im} d^X_{i+1} \to X_{i-1}) \cong H_i(A) \quad \text{and} \quad \operatorname{coker} \bar{d}^X_i \cong \operatorname{coker}(X_i \to \ker d^X_{i-1}) \cong H_{i-1}(A),$$

and the morphisms induced by f_i and g_i (resp., f_{i-1} and g_{i-1}) on the first (resp., second) of these are just the morphisms f_i^* and g_i^* (resp., f_{i-1}^* and g_{i-1}^*). Our exact sequence then becomes

$$H_i(A) \xrightarrow{f_i^*} H_i(B) \xrightarrow{g_i^*} H_i(C) \xrightarrow{\delta_i} H_{i-1}(A) \xrightarrow{f_{i-1}^*} H_{i-1}(B) \xrightarrow{g_{i-1}^*} H_{i-1}(C).$$

Taken for all i, these yield the long exact sequence. □

EXAMPLE 12.3.14. Consider the complexes $A_\cdot$, $B_\cdot$, and $C_\cdot$ of abelian groups with $A_i = \mathbb{Z}/4\mathbb{Z}$, $B_i = \mathbb{Z}/8\mathbb{Z}$, and $C_i = \mathbb{Z}/2\mathbb{Z}$ for all i and $d^A_i = d^B_i = 0$ for all i and d^C_i equal to multiplication by 4 for all i. Then $H_i(A) = \mathbb{Z}/4\mathbb{Z}$, $H_i(B) = 2\mathbb{Z}/4\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z}$, and $H_i(C) = \mathbb{Z}/2\mathbb{Z}$ for all i. We have a short exact sequence

$$0 \to A_\cdot \xrightarrow{f_\cdot} B_\cdot \xrightarrow{g_\cdot} C_\cdot \to 0$$

for maps $f_\cdot \colon A_\cdot \to B_\cdot$ induced by multiplication by 2 on each term and $g_\cdot \colon B_\cdot \to C_\cdot$ given by reduction modulo 2 on each term. The resulting long exact sequence has the form

$$\cdots \xrightarrow{\delta} \mathbb{Z}/4\mathbb{Z} \twoheadrightarrow \mathbb{Z}/2\mathbb{Z} \xrightarrow{0} \mathbb{Z}/2\mathbb{Z} \xhookrightarrow{\delta} \mathbb{Z}/4\mathbb{Z} \twoheadrightarrow \mathbb{Z}/2\mathbb{Z} \xrightarrow{0} \mathbb{Z}/2\mathbb{Z} \xrightarrow{\delta} \cdots .$$

We briefly describe cochain complexes and cohomology, which simply amount to a change of indexing from decreasing to increasing.

DEFINITION 12.3.15. A *cochain complex* is a collection $A^\cdot = (A^i, d^i_A)_{i\in\mathbb{Z}}$ of objects A^i and morphisms $d^i_A \colon A^i \to A^{i+1}$ such that $d^i_A \circ d^{i-1}_A = 0$ for all $i \in \mathbb{Z}$. The morphism $d^i_A \colon A^i \to A^{i+1}$ is called the ith *differential* of the cochain complex $A^\cdot$.

We then have the notion of cohomology.

DEFINITION 12.3.16. The *ith cohomology* of a cochain complex $A^\cdot$ in an abelian category is the object

$$H^i(A) = \frac{\ker d^i_A}{\operatorname{im} d^{i-1}_A}.$$

REMARK 12.3.17. Much as with complexes, we can speak of morphisms of cochain complexes $f^\cdot : A^\cdot \to B^\cdot$, which are collections of morphisms $f^i : A^i \to B^i$ such that $d_B^i \circ f^i = f^{i+1} \circ d_A^i$ for all $i \in \mathbb{Z}$. Again, short exact sequences

$$0 \to A^\cdot \xrightarrow{f^\cdot} B^\cdot \xrightarrow{g^\cdot} C^\cdot \to 0$$

of cochain complexes give rise to long exact sequences in cohomology, but now of the form

$$\cdots \to H^i(A) \xrightarrow{f_*^i} H^i(B) \xrightarrow{g_*^i} H^i(C) \xrightarrow{\delta^i} H^{i+1}(A) \to \cdots$$

DEFINITION 12.3.18. Let $A_\cdot = (A_i, d_i^A)$ and $B_\cdot = (B_i, d_i^B)$ be chain complexes. Let $f_\cdot, g_\cdot : A_\cdot \to B_\cdot$ be morphisms of chain complexes.

a. A *chain homotopy* from $f_\cdot$ to $g_\cdot$ is a sequence $s_\cdot = (s_i)_{i\in\mathbb{Z}}$ of morphisms $s_i : A_i \to B_{i+1}$ satisfying

$$f_i - g_i = d_{i+1}^B \circ s_i + s_{i-1} \circ d_i^A$$

for all $i \in \mathbb{Z}$.

b. We say that $f_\cdot$ and $g_\cdot$ are *chain homotopic*, and write $f_\cdot \sim g_\cdot$, if there exists a homotopy from $f_\cdot$ to $g_\cdot$.

c. If $f_\cdot$ is (chain) homotopic to 0, then $f_\cdot$ is said to be *null-homotopic*.

REMARK 12.3.19. For cochain complexes $A^\cdot$ and $B^\cdot$ and morphisms $f^\cdot, g^\cdot : A^\cdot \to B^\cdot$, a chain homotopy from $f^\cdot$ to $g^\cdot$ is a sequence $s^\cdot = (s^i)_{i\in\mathbb{Z}}$ of morphisms $s^i : A^i \to B^{i-1}$ such that $f^i - g^i = d_B^{i-1} \circ s^i + s^{i+1} \circ d_A^i$.

The morphisms $s_\cdot$ defining a null-homotopy fit into a diagram

$$\begin{tikzcd}
\cdots \arrow[r] & A_{i+1} \arrow[r, "d_{i+1}^A"] \arrow[d, "f_{i+1}"'] & A_i \arrow[r, "d_i^A"] \arrow[d, "f^i"'] \arrow[dl, "s_i"'] & A_{i-1} \arrow[r] \arrow[d, "f_{i-1}"] \arrow[dl, "s_{i-1}"'] & \cdots \\
\cdots \arrow[r] & B_{i+1} \arrow[r, "d_{i+1}^B"] & B_i \arrow[r, "d_i^B"] & B_{i-1} \arrow[r] & \cdots .
\end{tikzcd}$$

PROPOSITION 12.3.20. *Assume that $f_\cdot$ and $g_\cdot$ are chain homotopic morphisms $A_\cdot \to B_\cdot$. Then the morphisms f_i^* and g_i^* on homology are equal for all $i \in \mathbb{Z}$.*

PROOF. It suffices to assume that $g = 0$, since the ith cohomology functor from $\mathbf{Ch}(\mathscr{C})$ to $\mathscr{C}$ is additive. So, we must show that $f_i^* = 0$ for all i, which is to say that $f_i(\ker d_i^A) \subseteq \operatorname{im} d_{i-1}^A$. Since $f_i = d_{i+1}^B \circ s_i + s_{i-1} \circ d_i^A$, we have

$$f^i(\ker d_i^A) = d_{i+1}^B(s_i(\ker d_i^A)) \subseteq \operatorname{im} d_{i+1}^B,$$

so $f_i^* = 0$. □

DEFINITION 12.3.21. A morphism of complexes $f_\cdot : A_\cdot \to B_\cdot$ is a *homotopy equivalence* if there exists a morphism $g_\cdot : B_\cdot \to A_\cdot$ such that $g_\cdot \circ f_\cdot \sim \mathrm{id}_{A_\cdot}$ and $f_\cdot \circ g_\cdot \sim \mathrm{id}_{B_\cdot}$.

12.4. Projective and injective objects

We continue to work in an abelian category $\mathscr{C}$.

DEFINITION 12.4.1.

a. We say that an epimorphism $g\colon B \to C$ is *split* if there exists a morphism $t\colon C \to B$ with $g \circ t = \mathrm{id}_C$. In this case, we say that t is a splitting of g.

b. We say that a monomorphism $f\colon A \to B$ is *split* if there exists a morphism $s\colon B \to A$ with $s \circ f = \mathrm{id}_A$. In this case, we say that s is a splitting of f.

c. We say that a short exact sequence

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0 \tag{12.4.1}$$

splits if there exists an isomorphism $w\colon A \oplus C \xrightarrow{\sim} B$ with $w(a,0) = f(a)$ and $g(w(0,c)) = c$ for all $a \in A$ and $c \in C$.

EXAMPLE 12.4.2. The exact sequence of abelian groups

$$0 \to \mathbb{Z}/3\mathbb{Z} \xrightarrow{\cdot 2} \mathbb{Z}/6\mathbb{Z} \xrightarrow{\text{mod } 2} \mathbb{Z}/2\mathbb{Z} \to 0$$

is split, but

$$0 \to \mathbb{Z}/2\mathbb{Z} \xrightarrow{\cdot 2} \mathbb{Z}/4\mathbb{Z} \xrightarrow{\text{mod } 2} \mathbb{Z}/2\mathbb{Z} \to 0$$

is not.

PROPOSITION 12.4.3. *The following conditions on a short exact sequence*

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

are equivalent:

i. The sequence splits.

ii. The monomorphism $f\colon A \to B$ splits.

iii. The epimorphism $g\colon B \to C$ splits.

PROOF. We prove this in the category of R-modules.

(iii) ⇒ (ii): Suppose we have a splitting map $t\colon C \to B$. Then define $s\colon B \to A$ by $s(b) = a$ where $f(a) = b - t(g(b))$. This is well-defined as f is injective, and such an a exists since

$$g(b - t(g(b))) = g(b) - g(t(g(b))) = g(b) - g(b) = 0.$$

It splits f as

$$s(f(a)) = s(b) - s(t(g(b))) = s(b),$$

the latter step using the fact that $s \circ t = 0$, which follows in turn from

$$f(s(t(c))) = t(c) - t(g(t(c))) = t(c) - t(c) = 0.$$

(ii) ⇒ (iii): Suppose that we have a splitting map $s\colon B \to A$. Then define $t\colon C \to B$ by $t(c) = b - f(s(b))$ where $g(b) = c$. To see this is well defined, note that $f(a) - f(s(f(a))) = 0$ for any $a \in A$.

(ii)+(iii) $\Rightarrow$ (i): Define $w(a,c) = f(a)+t(c)$. Its inverse is $w'(b) = (s(b),g(b))$. To see this, we check that

$$w' \circ w(a,c) = (s(f(a)+t(c)), g(f(a)+t(c))) = (s(f(a)), g(t(c))) = (a,c)$$

for $a \in A$ and $c \in C$, and note that

$$w \circ w'(b) = w(s(b),g(b)) = f(s(b))+t(g(b)).$$

for $b \in B$. Set $c = g(b)$, so that $g(b-t(c)) = 0$, and let $a \in A$ be such that $f(a) = b-t(c)$. Then

$$(f \circ s + t \circ g)(b) = f(s(f(a))) + f(s(t(c))) + t(g(f(a))) + t(g(t(c))) = f(a)+t(c) = b.$$

(i) $\Rightarrow$ (iii) Set $t(c) = w(0,c)$. Then $g(t(c)) = g(w(0,c)) = c$.

□

DEFINITION 12.4.4. An object P of an abelian category is *projective* if for epimorphism $\pi\colon B \to C$ and every morphism $g\colon P \to C$, there exists an morphism $f\colon P \to B$ such that $g = \pi \circ f$.

REMARK 12.4.5. The property of P being projective is represented by the existence of g in the commutative diagram

$$\begin{tikzcd} & P \arrow[dl, dashed, "f"'] \arrow[d, "g"] & \\ B \arrow[r, "\pi"] & C \arrow[r] & 0 \end{tikzcd}$$

with exact lower row.

PROPOSITION 12.4.6. *Free R-modules are projective.*

PROOF. Let F be a free R-module with basis X. Let $\pi\colon B \to C$ be a surjective R-module homomorphism, and suppose that $g\colon F \to C$ is an R-module homomorphism. For each $x \in X$, let b be an element of B such that $\pi(b) = g(x)$. Since F is free, we may defined $f\colon F \to B$ by $f(x) = b$ for each $x \in X$. Then $\pi(f(x)) = g(x)$ for all $x \in X$, so $\pi \circ f = g$ as X generates F. □

EXAMPLE 12.4.7. Not every projective module need be free. For example, consider $R = \mathbb{Z}/6\mathbb{Z}$. We claim that $P = \mathbb{Z}/3\mathbb{Z}$ is a projective R-module. To see this, suppose that B is a $\mathbb{Z}/6\mathbb{Z}$-module and $g\colon B \to \mathbb{Z}/3\mathbb{Z}$ is surjective. Take any $b \in B$ with $g(b) = 1$. Then the $\mathbb{Z}/6\mathbb{Z}$ submodule generated by b is isomorphic to $\mathbb{Z}/3\mathbb{Z}$, and hence $1 \mapsto b$ defines a splitting of g.

Note that P is not projective as a $\mathbb{Z}$-module (abelian group) since the quotient map $\mathbb{Z} \to \mathbb{Z}/3\mathbb{Z}$ does not split. In fact, every projective $\mathbb{Z}$-module is free.

We describe some equivalent conditions for projectivity.

PROPOSITION 12.4.8. *The following conditions on an R-module P are equivalent:*

i. P is projective,

ii. every surjection $\pi\colon M \to P$ of R-modules is split, and

iii. P is a direct summand of a free R-module.

PROOF. If P is projective and $\pi\colon M \to P$ is a surjection, then the identity map $\mathrm{id}_P\colon P \to P$ lifts to a homomorphism $f\colon P \to M$ such that $\mathrm{id}_P = \pi \circ f$, so (i) implies (ii). If (ii) holds, then choose a set of generators X of P, and let F be the free R-module on X, which comes equipped with a surjection $\pi\colon F \to P$ that restricts to id_X. This surjection is split, so P is a direct summand of F by Proposition 12.4.3, and therefore (iii) holds.

If P is a direct summand of a free module F with complement Q, then $\pi\colon B \to C$ is a surjection, and $g\colon P \to C$ is a homomorphism of R-modules, then we can extend g to $\tilde{g}\colon F \to C$ by setting $\tilde{g}(q) = 0$ for all $q \in Q$. We then have $\tilde{f}\colon F \to B$ such that $\pi \circ \tilde{f} = \tilde{g}$ by the projectivity of F, and the restriction $f = \tilde{f}|_P$ satisfies $\pi \circ f = g$. Thus, (iii) implies (i). □

EXAMPLE 12.4.9. For $n \geq 1$, the left $M_n(R)$-module L of column vectors under left multiplication is a direct summand of $M_n(R)$, which is isomorphic to L^n as a left R-module. Hence, L is projective, though it is not free for $n \geq 2$.

In the case that R is a principal ideal domain, we have the following.

COROLLARY 12.4.10. *If R is a principal ideal domain, then every projective R-module is free.*

PROOF. By the classification of finitely generated modules over a principal ideal domain, it suffices for finitely generated R-modules to show that any R-module of the form

$$A = R/(a_1) \oplus R/(a_2) \oplus \cdots R/(a_n)$$

for nonzero and nonunit $a_1, a_2, \dots, a_n \in R$ is not projective. Consider the obvious quotient map $R^n \to A$. That it splits means that each $R \to R/(a_i)$ splits. Then $R \cong (a_i) \oplus (x)$ for some $x \in R$, which means that R is free of rank 2 over itself, which is impossible (e.g., by the classification theorem).

The general case is left as an exercise. □

DEFINITION 12.4.11. An object I in an abelian category is *injective* if for every monomorphism $\iota\colon A \to B$ and every morphism $f\colon A \to I$, there exists a morphism $g\colon B \to I$ such that $f = g \circ \iota$.

REMARK 12.4.12. The property of I being injective is represented by the existence of g in the commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & A \arrow[r, "\iota"] \arrow[d, "f"'] & B \arrow[dl, dashed, "g"] \\
 & I &
\end{tikzcd}$$

with exact upper row.

Dually to the analogous result projective modules, we have the following.

LEMMA 12.4.13. *As R-module I is injective if and only if every monomorphism $\iota\colon I \to R$ is split.*

We also have the following interesting criterion, which employs Zorn's lemma.

PROPOSITION 12.4.14 (Baer's criterion). *A left R-module I is injective if and only if every homomorphism $J \to I$ with J a left ideal of R may be extended to a map $R \to I$.*

PROOF. Let A be an R-submodule of an R-module B and $f\colon A \to I$ be an R-module homomorphism. It suffices to show that we can extend f to $g\colon B \to I$ with $g|_A = f$. Let X be the set of pairs (C,h) with C an R-submodule of B containing A and $h\colon C \to I$ an R-module homomorphism. We have a partial ordering on X given by $(C,h) < (C',h')$ if C is contained in C' and $h'|_C = h$. Given a chain $\mathscr{C}$ in X, we have an upper bound (C,h) with $C = \bigcup_{(D,k)\in\mathscr{C}} D$ such that if $d \in D$ for $(D,k) \in \mathbb{C}$, then $h(d) = k(d)$. By Zorn's lemma, $\mathscr{C}$ has a maximal element (M,l).

Suppose first that $M \neq B$, and let $b \in B - M$. Consider the left ideal

$$J = \{r \in R \mid rb \in M\}$$

of R. Define $s\colon J \to I$ by $s(r) = l(rb)$ for $r \in J$. This R-module homomorphism may be extended to $t\colon R \to I$ by assumption. Then define $N = M + Rb$ and let $q\colon N \to I$ be the unique R-module homomorphism such that $q|_M = l|_M$ and $q(rb) = t(r)$ for all $r \in R$. This exists as $M \cap Rb = Jb$, and $l(rb) = s(r) = t(r)$ for $r \in J$. (Also, if $rb = 0$, then $r \in J$, so $q(rb) = l(rb) = 0$ in this instance.) The existence of q gives a contradiction of the maximality of M. Thus $M = B$, and we are done. □

EXAMPLE 12.4.15.

a. $\mathbb{Q}$ is an injective $\mathbb{Z}$-module.

b. $\mathbb{Z}/n\mathbb{Z}$ is an injective $\mathbb{Z}/n\mathbb{Z}$-module for any $n \geq 1$.

c. $\mathbb{Z}/3\mathbb{Z}$ is an injective $\mathbb{Z}/6\mathbb{Z}$-module, but not an injective $\mathbb{Z}/9\mathbb{Z}$-module.

We have a very nice description of injective objects in **Ab**.

DEFINITION 12.4.16. An abelian group D is called *divisible* if multiplication by n is surjective on D for every natural number n.

PROPOSITION 12.4.17. *An abelian group is injective if and only if it is divisible.*

PROOF. Let D be injective, and take $d \in D$. Then there exists a group homomorphism $\phi\colon \mathbb{Z} \to D$ with $1 \mapsto d$. We also have the multiplication-by-n map on $\mathbb{Z}$, which is injective. By injectivity of D, we have a map $\theta\colon \mathbb{Z} \to D$ with $\phi = n\theta$. Then $d = n\theta(1)$, so D is divisible.

Conversely, let D be divisible. By Baer's criterion, it suffices to show that every homomorphism $\phi\colon n\mathbb{Z} \to D$ with $n \geq 1$ extends to a homomorphism $\theta\colon \mathbb{Z} \to D$. Such a ϕ is determined by $d = \phi(n)$. Let $d' \in D$ be such that $nd' = d$. Set $\theta(1) = d'$. □

12.5. Exact functors

Despite the fact that additive functors preserve direct sums, they may not preserve exact sequences. We make the following definitions.

DEFINITION 12.5.1. Let $F\colon \mathscr{C} \to \mathscr{D}$ be an additive functor of abelian categories.

a. We say that F is *left exact* if for every left short exact sequence $0 \to A \to B \to C$ in $\mathscr{C}$, the sequence $0 \to F(A) \to F(B) \to F(C)$ is exact in $\mathscr{D}$.

b. We say that F is *right exact* if for every right short exact sequence $A \to B \to C \to 0$ in $\mathscr{C}$, the sequence $F(A) \to F(B) \to F(C) \to 0$ is exact in $\mathscr{D}$.

c. We say that F is an *exact functor* if for every short exact sequence $0 \to A \to B \to C \to 0$ in $\mathscr{C}$, the sequence $0 \to F(A) \to F(B) \to F(C) \to 0$ is exact in $\mathscr{D}$.

REMARK 12.5.2. A contravariant additive functor $F\colon \mathscr{C} \to \mathscr{D}$ is left exact if the resulting covariant functor $\mathscr{C}^{\mathrm{op}} \to \mathscr{D}$ is left exact.

EXAMPLE 12.5.3. The functor $F\colon \mathbf{Ab} \to \mathbf{Ab}$ by $F(A) = A \oplus A$ with $F(f) = f \oplus f$ is exact.

TERMINOLOGY 12.5.4. We (somewhat loosely) say a functor *a* certain structure if its takes structures of one sort in a given category (induced from the source category) to those of the same sort in another (induced from the target category). For instance, exact functors are additive functors that preserve short exact sequences.

LEMMA 12.5.5. *Let $F\colon \mathscr{C} \to \mathscr{D}$ be an additive functor of abelian categories. The following are equivalent:*

i. F is exact,

ii. F is both left and right exact,

iii. F preserves all three-term exact sequences $A \to B \to C$, and

iv. F preserves all exact sequences.

PROOF. It is immediate that (iv) implies the other statements and also that (ii) implies (i).

Suppose that F that preserves all three-term exact sequences. To say that

$$0 \to A \to B \to C \to 0$$

is short exact is equivalent to saying that the three three-term sequences $0 \to A \to B$, $A \to B \to C$, and $B \to C \to 0$ are all exact. Since exactness of these is preserved by F, so is exactness of the original short exact sequence. So, (iii) implies (i).

If F is an exact functor, take any exact sequence $A_{\cdot}$. Then $0 \to \ker d_i \to A_i \to \operatorname{im} d_i \to 0$ is short exact, so

$$0 \to F(\ker d_i) \to F(A_i) \xrightarrow{F(d_i)} F(\operatorname{im} d_i) \to 0$$

is exact as well. It $F(d_i)$ follows that image $F(\operatorname{im} d_i)$ and kernel $F(\ker d_i)$ for all i, so we have

$$\operatorname{im} F(d_i) = F(\operatorname{im} d_i) = F(\ker d_{i-1}) = \ker F(d_{i-1}).$$

Thus, $F(A_{\cdot})$ is exact. Thus, (i) implies (iv), which finishes the proof. □

REMARK 12.5.6. The reader may also check that an additive functor of abelian categories is left (resp., right) exact if and only if it sends short exact sequences to left (resp., right) short exact sequences.

Recall that for an additive category $\mathscr{C}$, the functors h^X (and h_X) may be viewed as taking values in $\mathbf{Ab}$, and clearly such functors are additive. In fact, they are also left exact.

LEMMA 12.5.7. *Let $\mathscr{C}$ be an abelian category, and let X be an object of $\mathscr{C}$.*

a. The functor $h_X \colon \mathscr{C} \to \mathbf{Ab}$ is left exact.

b. The functor $h^X \colon \mathscr{C}^{\mathrm{op}} \to \mathbf{Ab}$ is left exact.

PROOF. Let

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

be an exact sequence in $\mathscr{C}$. Applying h_X, we obtain homomorphisms

$$0 \to \mathrm{Hom}_{\mathscr{C}}(X,A) \xrightarrow{h_X(f)} \mathrm{Hom}_{\mathscr{C}}(X,B) \xrightarrow{h_X(g)} \mathrm{Hom}_{\mathscr{C}}(X,C)$$

of abelian groups, and we claim this sequence is exact. If $h_X(f)(\alpha) = 0$, then $f \circ \alpha = 0$, but f is a monomorphism, so $\alpha = 0$. Since h_X is a functor, we have $h_X(g) \circ h_X(f) = 0$, and if $\beta \in \ker h_X(g)$, then $g \circ \beta = 0$. Naturality of the kernel implies that β factors through a morphism $X \to \ker g$. But we have canonical isomorphisms

$$A \xrightarrow{\sim} \mathrm{coim}\, f \xrightarrow{\sim} \mathrm{im}\, f \xrightarrow{\sim} \ker g,$$

the first as f is a monomorphism, and the composite of the composite of these with the canonical morphism $\ker g \to B$ is g. Therefore, we obtain a morphism $\alpha \colon X \to A$ satisfying $f \circ \alpha = g$. This proves part a, and part b is just part a with $\mathscr{C}$ replaced by $\mathscr{C}^{\mathrm{op}}$. □

LEMMA 12.5.8. *Let R be a ring, and let N be a right R-module. The tensor product functor $t_N \colon R$-$\mathbf{mod} \to \mathbf{Ab}$ given on objects by $t_N(M) = N \otimes_R M$ and on morphisms by $t_N(g) = \mathrm{id}_N \otimes g$ is right exact.*

PROOF. Since tensor products commute with direct sums, t_N is additive. Let

$$A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

be a right short exact sequence of R-modules. The group $N \otimes_R C$ is generated by simple tensors $n \otimes c$ with $n \in N$ and $c \in C$ and

$$n \otimes c = n \otimes g(b) = (\mathrm{id}_N \otimes g)(n \otimes b)$$

for any $b \in B$ with $g(b) = c$, we have that $t_N(g)$ is surjective. We need then only define an inverse to the surjection $\bar{g} \colon \operatorname{coker} t_N(f) \to N \otimes_R C$. For this, we consider the map $\theta \colon N \times C \to \operatorname{coker} t_N(f)$ given on $(n,c) \in N \times C$ by picking $b \in B$ with $g(b) = c$ and then setting $\theta(n,c) = n \otimes b + \mathrm{im}\, t_N(f)$. If $g(b') = c$, then $g(b - b') = 0$, so $b - b' = f(a)$ for some $a \in A$, and then $n \otimes (b - b') = t_N(f)(n \otimes a)$, so θ is well-defined and then easily seen to be biadditive and R-balanced. The induced map $\Theta \colon N \otimes_R C \to \operatorname{coker} t_N(f)$ is inverse to $\bar{g}$ by definition. □

LEMMA 12.5.9. *Let $\mathscr{C}$ be an abelian category. A sequence*

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C$$

is exact if every sequence

$$0 \to \mathrm{Hom}_{\mathscr{C}}(X,A) \xrightarrow{h_X(f)} \mathrm{Hom}_{\mathscr{C}}(X,B) \xrightarrow{h_X(g)} \mathrm{Hom}_{\mathscr{C}}(X,C)$$

is exact.

PROOF. For $X = A$, we get

$$g \circ f = h_X(g) \circ h_X(f)(\mathrm{id}_A) = 0,$$

so we have a monomorphism $s\colon \operatorname{im} f \to \ker g$. For $X = \ker g$ and $\beta\colon \ker g \to B$ the natural monomorphism defined by the kernel, we have $h^X(g)(\beta) = g \circ \beta = 0$, so there exists $\alpha\colon \ker g \to A$ with $f \circ \alpha = \beta$. We then have that β factors a morphism $t\colon \ker g \to \operatorname{im} f$ inverse to s. □

PROPOSITION 12.5.10. *Any left (resp., right) adjoint to ia functor between abelian categories is left (resp., right) exact.*

PROOF. We treat the case of left exactness, the other case simply being the corresponding statement in opposite categories. , Let $G\colon \mathscr{C} \to \mathscr{D}$ be an additive functor of abelian categories that admits a left adjoint F. Suppose that

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C$$

is a left exact sequence in $\mathscr{C}$. Then for any $D \in \mathrm{Obj}(\mathscr{D})$, the sequence

$$0 \to h_{F(D)}(A) \xrightarrow{h_{F(D)}(f)} h_{F(D)}(B) \xrightarrow{h_{F(D)}(g)} h_{F(D)}(C)$$

is left exact. Since F is left adjoint to G, this sequence is isomorphic to

$$0 \to h_D(G(A)) \xrightarrow{h_D(G(f))} h_D(G(B)) \xrightarrow{h_D(G(g))} h_D(G(C))$$

as a sequence of abelian groups. Since this holds for all D, the sequence

$$0 \to G(A) \xrightarrow{G(f)} G(B) \xrightarrow{G(g)} G(C)$$

is exact. □

PROPOSITION 12.5.11. *Let R be a ring, and fix an R-module M.*

*a. The covariant homomorphism functor $h_M\colon R$-**mod** $\to$ **Ab** is exact if and only if M is R-projective.*

*b. The contravariant homomorphism functor $h^M\colon R$-**mod** $\to$ **Ab** is exact if and only if M is R-injective.*

PROOF. We prove part a. Suppose that the functor is exact. Then for any epimorphism $g\colon B \to P$ we have an epimorphism

$$\mathrm{Hom}_R(P,B) \to \mathrm{Hom}_R(P,P),$$

and any inverse image t of id_P is the desired splitting map of g.

On the other hand, suppose that P is projective. Consider an exact sequence

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0.$$

Then we have a diagram

$$\mathrm{Hom}_R(P,A) \xrightarrow{h_P(f)} \mathrm{Hom}_R(P,B) \xrightarrow{h_P(g)} \mathrm{Hom}_R(P,C) \to 0.$$

That this is a complex is immediate. Surjectivity of $h_P(g)$ follows immediately from the definition of a projective module. Finally, let $h \in \ker h_P(g)$, so $h\colon P \to \ker g$. Then $A \to \ker g$ is an epimorphism, and we have by projectivity of P a map $j\colon P \to A$ with with $f \circ j = h$, i.e., $h_P(f)(j) = h$. □

REMARK 12.5.12. If R is a commutative ring, then $\mathrm{Hom}_R(A,B)$ for R-modules A and B may be viewed as an R-module under $(r \cdot f)(a) = r \cdot f(a)$. It follows easily that $\mathrm{Hom}_R(A, \cdot)$ is an additive functor from the category of R-modules to itself which is exact if A is projective.

The following embedding theorem, the proof of which is beyond the scope of these notes, allows us to do most of the homological algebra that can be done in the category of R-modules for any R in an arbitrary abelian category.

THEOREM 12.5.13 (Freyd-Mitchell). *If $\mathscr{C}$ is a small abelian category, then there exists a ring R and an exact, fully faithful functor $\mathscr{C} \to R$-**mod**.*

In other words, $\mathscr{C}$ is equivalent to a full, abelian subcategory of R-**mod** for some ring R. We can use this as follows: suppose there is a result we can prove about exact diagrams in R-modules for all R, like the snake lemma. We then have the result in all abelian categories, since we can take a small full, abelian subcategory containing the objects in which we are interested and embed it into some category of left R-modules. If the result holds in that category, then by exactness of the embedding, the result will hold in the original category.

12.6. Projective and injective resolutions

DEFINITION 12.6.1. Let $\mathscr{C}$ be an abelian category, and let A be an object in $\mathscr{C}$.

a. A *resolution* of A is a complex $C_\cdot$ of objects in $\mathscr{C}$ together with an *augmentation morphism* $\varepsilon^C\colon C_0 \to A$ such that the *augmented complex*

$$\cdots \to C_1 \xrightarrow{d_1^C} C_0 \xrightarrow{\varepsilon^C} A \to 0$$

is exact.

b. A *projective resolution* of A is a resolution of A by a complex of projective objects.

DEFINITION 12.6.2. An abelian category $\mathscr{C}$ is said to have *sufficiently many* (or *enough*) projectives if for every $A \in \mathrm{Obj}(\mathscr{C})$, there exists a projective object $P \in \mathrm{Obj}(\mathscr{C})$ and an epimorphism $P \to A$.

Since free modules are projective, R-**mod** has enough projectives.

REMARK 12.6.3. If $\mathscr{C}$ has enough projectives, then every object in $\mathscr{C}$ has a projective resolution. (We leave the proof as an exercise.)

EXAMPLES 12.6.4. We have the following examples of projective resolutions, all of which are in fact resolutions by free modules:

a. In **Ab**, the abelian group $\mathbb{Z}/n\mathbb{Z}$ has a projective resolution

$$0 \to \mathbb{Z} \xrightarrow{\cdot n} \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z} \to 0.$$

b. Consider $R = \mathbb{Z}[X]$, and let $A = \mathbb{Z}[X]/(n, X^2+1)$. Then we have a projective resolution

$$0 \to \mathbb{Z}[X] \xrightarrow{(\cdot(X^2+1),\cdot n)} \mathbb{Z}[X] \oplus \mathbb{Z}[X] \xrightarrow{(\cdot n)-(\cdot(X^2+1))} \mathbb{Z}[X] \to \mathbb{Z}[X]/(n, X^2+1) \to 0.$$

c. Consider the ring $R = \mathbb{Z}[X]/(X^n - 1)$. (This is isomorphic to the *group ring* $\mathbb{Z}[\mathbb{Z}/n\mathbb{Z}]$.) We have a projective resolution of $\mathbb{Z}$:

$$\cdots \to R \xrightarrow{N} R \xrightarrow{X-1} R \xrightarrow{N} R \xrightarrow{X-1} R \to \mathbb{Z} \to 0,$$

where $N = \sum_{i=0}^{n-1} X^i$.

PROPOSITION 12.6.5. *Let $P. \to A$ and $Q. \to B$ be projective resolutions in an abelian category, and suppose that $g\colon A \to B$ is an R-module homomorphism. Then g extends to a morphism $f.\colon P. \to Q.$ of chain complexes such that*

$$\begin{tikzcd}
\cdots \arrow[r] & P_2 \arrow[r] \arrow[d, "f_2"] & P_1 \arrow[r] \arrow[d, "f_1"] & P_0 \arrow[r] \arrow[d, "f_0"] & A \arrow[r] \arrow[d, "g"] & 0 \\
\cdots \arrow[r] & Q_2 \arrow[r] & Q_1 \arrow[r] & Q_0 \arrow[r] & B \arrow[r] & 0
\end{tikzcd}$$

commutes. Furthermore, any other lift of g is chain homotopic to $f.$.

PROOF. Let $P = (P_i, d_i)$ and $Q = (Q_i, d_i')$, and let ε and ε' denote the respective augmentation maps. Then $g \circ \varepsilon\colon P_0 \to B$. Since ε' is an epimorphism, we have a map $f_0\colon P_0 \to Q_0$ lifting $g \circ \varepsilon$. Now f_0 induces a map

$$\bar{f}_0\colon \ker\varepsilon \to \ker\varepsilon',$$

and since $\operatorname{im} d_1 = \ker\varepsilon$ and $\ker d_1' = \ker\varepsilon'$, we have an epimorphism $Q_1 \to \ker d_1'$, and we again use projectivity, this time of Q_1, to lift $\bar{f}_0 \circ d_1$ to a map f_1 as in the diagram. We continue in this manner to obtain $f.$.

Now, for uniqueness up to chain homotopy, it suffices to show that if $g = 0$, then $f.$ is chain homotopic to zero. Well, $d_0' \circ f_0 = g \circ d_0 = 0$, so $f_0(P_0) \subseteq \operatorname{im} d_1'$. By projectivity of P_0, we have $s_0\colon P_0 \to Q_1$ with

$$f_0 = d_1' \circ s_0 + s_{-1} \circ d_0 = d_1' \circ s_0,$$

where we have set $s_i = 0$ for $i < 0$ (and $d_i = 0$ for $i \le 0$). Now $h_1 = f_1 - s_0 \circ d_1$ satisfies

$$d_1' \circ h_1 = d_1' \circ f_1 - d_1' \circ s_0 \circ d_1 = f_0 \circ d_1 - f_0 \circ d_1 = 0,$$

so $\operatorname{im} h_1 \subseteq \operatorname{im} d_2'$. Thus, we have $s_1\colon P_1 \to Q_1$ lifting h_1, i.e., so that

$$d_2' \circ s_1 = h_1 = f_1 - s_0 \circ d_1,$$

as desired. We continue in this fashion to obtain all s_i. □

PROPOSITION 12.6.6 (Horseshoe lemma). *Suppose that*

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

is a short exact sequence in an abelian category and that $(P^A_{\cdot}, \varepsilon^A)$ and $(P^C_{\cdot}, \varepsilon^C)$ are projective resolutions of A and C respectively. Then there exists a projective resolution $(P^B_{\cdot}, \varepsilon^B)$ of B with $P_i^B = P_i^A \oplus P_i^C$ for each i and such that the diagram

(12.6.1)

$$\begin{tikzcd}
0 \arrow[r] & P^A_{\cdot} \arrow[r, "\iota_{\cdot}"] \arrow[d, "\varepsilon^A"] & P^B_{\cdot} \arrow[r, "p_{\cdot}"] \arrow[d, "\varepsilon^B"] & P^C_{\cdot} \arrow[r] \arrow[d, "\varepsilon^C"] & 0 \\
0 \arrow[r] & A \arrow[r, "f"] \arrow[d] & B \arrow[r, "g"] \arrow[d] & C \arrow[r] \arrow[d] & 0 \\
& 0 & 0 & 0 &
\end{tikzcd}$$

commutes, where $\iota_{\cdot}$ and $p_{\cdot}$ are the natural maps on each term.

PROOF. Choose a lift t_0 of ε^C to $P_0^C \to B$, and let

$$\varepsilon^B = f \circ \varepsilon^A + t_0 \circ p_0.$$

Then ε^B is clearly surjective, and we have the desired commutativity of the "first two" squares. Next, letting $d^X_{\cdot}$ denote the boundary maps with $X = A$, C, we may define the boundary map d_1^B for B similarly. That is, consider a lift of $d_1^C \colon P_1^C \to \ker \varepsilon^C$ to a map $t_1 \colon P_1^C \to \ker \varepsilon^B$, and define

$$d_1^B = \iota_0 \circ d_1^A + t_1 \circ p_1.$$

Then d_1^B maps onto $\ker d_1^B$ and makes the next two squares commute. We then continue in this fashion. □

LEMMA 12.6.7. *Let $\mathscr{C}$ be an abelian category and $P_{\cdot}$ a split long exact sequence of projectives with $P_i = 0$ for $i < 0$. Then $P_{\cdot}$ is a projective object in $\mathbf{Ch}(\mathscr{C})$.*

PROOF. Let $P_{\cdot}$ be a split exact sequence of projectives in $\mathscr{C}$. In other words, we may write each $P_0 = Q_0$ and $P_i = Q_i \oplus Q_{i-1}$ for $i \geq 1$, where Q_i is a projective object in $\mathscr{C}$, and the morphism $P_i \to P_{i-1}$ is simply the composition of the projection $P_i \to Q_{i-1}$ with the inclusion $Q_{i-1} \to P_i$. Suppose that $\pi_{\cdot} \colon A_{\cdot} \to P_{\cdot}$ is a epimorphism of complexes. Since Q_i is projective, there exists a splitting $s_i \colon Q_i \to A_i$ of the composition of π_i with projection to Q_i. Then

$$t_i = s_i \oplus s_{i-1} \colon P_i = Q_i \oplus Q_{i-1} \to A_i$$

is a splitting of π_i. Since $t_{\cdot}$ is a morphism of complexes, it is a splitting of $\pi_{\cdot}$. □

REMARK 12.6.8. Every split exact complex is the cone of a complex with zero differentials.

REMARK 12.6.9. Though we shall not prove it, every projective object in the category of chain complexes over an abelian category is a split exact sequence of projectives. Also, the projective objects in the category of bounded below chain complexes (or those in nonnegative degrees) are the bounded below exact sequences of projectives (which automatically split).

DEFINITION 12.6.10. We say that a functor $F \colon \mathscr{C} \to \mathscr{D}$ between categories *preserves projectives* if it takes projective objects in $\mathscr{C}$ to projective objects in $\mathscr{D}$.

PROPOSITION 12.6.11. *Let $\mathscr{C}$ and $\mathscr{D}$ be an abelian category. Let $F\colon \mathscr{C} \to \mathscr{D}$ be a functor that is left adjoint to an exact functor $G\colon \mathscr{D} \to \mathscr{C}$. Then F preserves projectives.*

PROOF. Let P be a projective object in $\mathscr{C}$. Let $f\colon A \to B$ be a epimorphism in $\mathscr{D}$. We must show that $h_{F(P)}(f)\colon F(A) \to F(B)$ is an epimorphism. Note that we have a commutative diagram

$$\begin{tikzcd}
\mathrm{Hom}_{\mathscr{C}}(P,G(A)) \arrow[r, "h_P(G(f))"] \arrow[d, "\wr"] & \mathrm{Hom}_{\mathscr{C}}(P,G(B)) \arrow[d, "\wr"] \\
\mathrm{Hom}_{\mathscr{D}}(F(P),A) \arrow[r, "h_{F(P)}(f)"] & \mathrm{Hom}_{\mathscr{D}}(F(P),B),
\end{tikzcd}$$

Exactness of G tells us that $G(f)$ is an epimorphism, and the upper horizontal map is then an epimorphism by the projectivity of P, hence the result. □

12.7. Derived functors

Suppose that $F\colon \mathscr{C} \to \mathscr{D}$ is a right exact functor between abelian categories $\mathscr{C}$ and $\mathscr{D}$ and that $\mathscr{C}$ has enough projectives. Then we could try to define the ith left derived functor of F (for $i \geq 0$) on an object A of $\mathscr{C}$ by $H_i(F(P.))$, where $P. \to A$ is a projective resolution of A. Of course, we must check that this definition is independent of the projective resolution chosen, and that we obtain induced maps on morphisms so that our map becomes a functor.

In the following, one may suppose that $\mathscr{C}$ is the category of R-**mod**, but it is often the case that $\mathscr{D}$ is some other category, like **Ab** or S-**mod** for some ring S.

PROPOSITION 12.7.1. *Let $F\colon \mathscr{C} \to \mathscr{D}$ be an additive functor between abelian categories $\mathscr{C}$ and $\mathscr{D}$. For $i \geq 0$, there are functors $L_iF\colon \mathscr{C} \to \mathscr{D}$ given on $A \in \mathscr{C}$ by $L_iF(A) = H_i(F(P.))$ for $P. \to A$ a projective resolution of $A \in \mathscr{C}$ and given on $g\colon A \to B$ in $\mathscr{C}$ by $L_iF(g) = F(f)_*\colon H_i(F(P.)) \to H_i(F(Q.))$ for $P. \to A$ and $Q. \to B$ projective resolutions and $f.\colon P. \to Q.$ a morphism of complexes $f\colon P. \to Q.$ compatible with the augmentations to A and B. These functors are dependent on the choices made up only to unique isomorphism.*

PROOF. The key point is that Proposition 12.6.5 tells us that $F(f)_*$ is independent of the choice of f, since any two choices are chain homotopic. In particular, given any two choices of $P.$ and $Q.$ of projective resolutions of A, and any choices of $f. \to P. \to Q.$ $f'_. \to Q. \to P.$ augmenting the identity morphism on A gives rise to morphisms on cohomology, the resulting maps $F(f)_*$ and $F(f')_*$ must be mutually inverse, since $f \circ f'$ and $f' \circ f$ augment the identity on A, as to the identity morphisms on $Q.$ and $P.$. Thus, $H_i(F(P.))$ is unique up to unique isomorphism, so L_iF is well-defined (up to unique isomorphism), and $L_iF(\mathrm{id}_A) = \mathrm{id}_{L_iF(A)}$. Similarly, it is easy to check that the uniqueness also implies that L_iF is compatible with compositions. □

DEFINITION 12.7.2. For an additive functor $F\colon \mathscr{C} \to \mathscr{D}$, the ith left derived functor $L_iF\colon \mathscr{C} \to \mathscr{D}$ of F is the functor defined by Proposition 12.7.1.

We have the following obvious corollaries of Lemma 12.7.1.

LEMMA 12.7.3. *Let $F\colon \mathscr{C} \to \mathscr{D}$ be a right exact functor of abelian categories. Then we have a canonical, natural isomorphism $L_0F \xrightarrow{\sim} F$ of functors.*

PROOF. Since F is right exact, the sequence

$$F(P_1) \to F(P_0) \to F(A) \to 0$$

is exact. Hence, we have

$$L_0F(A) = H_0(F(P.)) \cong F(A).$$

The reader will easily check the independence of the choice of resolution and naturality, as in Lemma 12.7.1. □

COROLLARY 12.7.4. *If P is a projective object, then $L_iF(P) = 0$ for $i \geq 1$.*

PROOF. Consider the projective resolution that is P in degree zero and 0 elsewhere, where the augmentation map $P \to P$ is the identity. This has the desired homology. □

Next, we prove that the L_iF are functors.

PROPOSITION 12.7.5. *To each morphism $f\colon A \to B$ in $\mathscr{C}$, we can associate morphisms*

$$L_iF(f)\colon L_iF(A) \to L_iF(B)$$

for all $i \geq 0$ in such a way that $L_iF\colon \mathscr{C} \to \mathscr{D}$ becomes a functor and $L_0F(f) = F(f)$. Furthermore, each L_iF is additive.

PROOF. The unique morphism $L_iF(f)$ is induced on homology by the morphism of chain complexes given in Proposition 12.6.5. Functoriality follows by canonicality of the map of homology.

To see additivity, note that $L_iF(0_A)$ is induced by the zero morphism of chain complexes and hence is is zero map on $L_iF(A)$. Similarly $L_iF(f+g)$, for $f,g\colon A \to B$, can be given by the sum of the induced maps on chain complexes, hence is given by the sum of the maps on homology. □

DEFINITION 12.7.6. For a right exact functor $F\colon \mathscr{C} \to \mathscr{D}$ of abelian categories, the functor L_iF is called ith *left derived functor* of F.

We see that L_0F and F are canonically naturally isomorphic functors.

DEFINITION 12.7.7. A homological δ-functor is a sequence of additive functors $F_i\colon \mathscr{C} \to \mathscr{D}$ for $i \in \mathbb{Z}$, together with, for every exact sequence

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

in $\mathscr{C}$, morphisms $\delta_i\colon F_i(C) \to F_{i-1}(A)$ fitting in a long exact sequence

$$\cdots \to F_i(A) \xrightarrow{F_i(f)} F_i(B) \xrightarrow{F_i(g)} F_i(C) \xrightarrow{\delta_i} F_{i-1}(A) \to \cdots$$

which are natural in the sense that if we have a morphism of short exact sequences in $\mathscr{C}$,

$$\begin{tikzcd}
0 \arrow[r] & A \arrow[r] \arrow[d] & B \arrow[r] \arrow[d] & C \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & A' \arrow[r] & B' \arrow[r] & C' \arrow[r] & 0,
\end{tikzcd}$$

then we obtain a morphism of long exact sequences in $\mathscr{D}$,

$$\begin{tikzcd}
\cdots \arrow[r] & F_i(A) \arrow[r] \arrow[d] & F_i(B) \arrow[r] \arrow[d] & F_i(C) \arrow[r] \arrow[d] & F_{i-1}(A) \arrow[r] \arrow[d] & \cdots \\
\cdots \arrow[r] & F_i(A') \arrow[r] & F_i(B') \arrow[r] & F_i(C') \arrow[r] & F_{i-1}(A') \arrow[r] & \cdots .
\end{tikzcd}$$

EXAMPLE 12.7.8. Define functors $F_0, F_1 \colon \mathbf{Ab} \to \mathbf{Ab}$ by $F_0(A) = A/pA$ and

$$F_1(A) = A[p] = \{a \in A \mid pa = 0\}$$

for any abelian group A, and set $F_i = 0$ otherwise. Given an exact sequence

$$0 \to A \to B \to C \to 0$$

in $\mathbf{Ab}$, we obtain a long exact sequence

$$0 \to A[p] \to B[p] \to C[p] \xrightarrow{\delta_1} A/pA \to B/pB \to C/pC \to 0$$

from the snake lemma applied to the diagram

$$\begin{tikzcd}
0 \arrow[r] & A \arrow[r] \arrow[d, "\cdot p"] & B \arrow[r] \arrow[d, "\cdot p"] & C \arrow[r] \arrow[d, "\cdot p"] & 0 \\
0 \arrow[r] & A \arrow[r] & B \arrow[r] & C \arrow[r] & 0.
\end{tikzcd}$$

This defines a δ-functor.

THEOREM 12.7.9. *For every short exact sequence*

$$0 \to A \to B \to C \to 0$$

in $\mathscr{C}$, there exist morphisms $\delta_i \colon L_iF(C) \to L_{i-1}F(A)$ such that the functors $L.F$ together with the maps $\delta.$ form a homological δ-functor.

PROOF. By the Horseshoe lemma, we have a projective resolution $P^X_\cdot \to X$ for $X = A, B, C$ fitting in a diagram (12.6.1). Now, applying F to the resolutions, we have split exact sequences

$$0 \to F(P_i^A) \to F(P_i^B) \to F(P_i^C) \to 0$$

for each i. The resulting exact sequence of complexes (which need not be split) yields a long exact sequence in homology

$$\cdots L_1F(B) \to L_1F(C) \xrightarrow{\delta_1} F(A) \to F(B) \to F(C) \to 0,$$

as desired.

It remains to check naturality. Consider a morphism of short exact sequences

$$\begin{tikzcd}
0 \arrow[r] & A \arrow[r, "f"] \arrow[d, "q^A"] & B \arrow[r, "g"] \arrow[d, "q^B"] & C \arrow[r] \arrow[d, "q^C"] & 0 \\
0 \arrow[r] & A' \arrow[r, "f'"] & B' \arrow[r, "g'"] & C' \arrow[r] & 0.
\end{tikzcd}$$

By Proposition 12.6.5, can extend q^A and q^C to maps of complexes $q^A_\cdot \colon P^A_\cdot \to P^{A'}_\cdot$ and $q^C_\cdot \colon P^C_\cdot \to P^{C'}_\cdot$. Suppose we have constructed $P^B_\cdot$ and $P^{B'}_\cdot$ via the Horseshoe lemma. We fit this all into a commutative diagram

(12.7.1)
$$\begin{tikzcd}
0 \arrow[rr] & & P^A_\cdot \arrow[rr, "\iota_\cdot"] \arrow[dr, "\varepsilon^A"'] \arrow[dd, "q^A_\cdot"] & & P^B_\cdot \arrow[rr, "p_\cdot"] \arrow[dr] & & P^C_\cdot \arrow[rr] \arrow[dr] \arrow[dd, "q^C_\cdot"] & & 0 & \\
 & 0 \arrow[rr] & & A \arrow[rr, "f"] \arrow[dd] & & B \arrow[rr, "g"] \arrow[dd] & & C \arrow[rr] \arrow[dd, "q^C"] & & 0 \\
0 \arrow[rr] & & P^{A'}_\cdot \arrow[rr, "\iota'"] \arrow[dr] & & P^{B'}_\cdot \arrow[rr, "p'"] \arrow[dr] & & P^{C'}_\cdot \arrow[rr] \arrow[dr] & & 0 & \\
 & 0 \arrow[rr] & & A' \arrow[rr, "f'"] & & B' \arrow[rr, "g'"] & & C' \arrow[rr] & & 0.
\end{tikzcd}$$

We also have splitting maps $j_i \colon P^C_i \to P^B_i$ and $k_i \colon P^B_i \to P^A_i$ for each i (and, similarly, maps j'_i and k'_i). For each X, let us denote the augmentation map by ε^X.

We must define a map $q^B_\cdot \colon P^B_\cdot \to P^{B'}_\cdot$ making the entire diagram (12.7.1) commute. We first note that

$$\begin{aligned} g' \circ (q^B \circ \varepsilon^B - \varepsilon^{B'} \circ j'_0 \circ q^C_0 \circ p_0) &= q^C \circ g \circ \varepsilon^B - \varepsilon^{C'} \circ p'_0 \circ j_0 \circ q^C_0 \circ p_0 \\ &= q^C \circ \varepsilon^C \circ p_0 - \varepsilon^{C'} \circ q^C_0 \circ p_0 = (q^C \circ \varepsilon^C - \varepsilon^{C'} \circ q^C_0) \circ p_0 = 0. \end{aligned}$$

Hence, there exists a map $\beta_0 \colon P^B_0 \to A'$ with

$$f' \circ \beta_0 = q^B \circ \varepsilon^B - \varepsilon^{B'} \circ j'_0 \circ q^C_0 \circ p_0.$$

Since $\varepsilon^{A'}$ is an epimorphism, we may choose $\alpha_0 \colon P^B_0 \to P^{A'}_0$ with $\varepsilon^{A'} \circ \alpha_0 = \beta_0$. Now set

$$q^B_0 = \iota'_0 \circ q^A_0 \circ k_0 + \iota'_0 \circ \alpha_0 \circ p_0 + j'_0 \circ q^C_0 \circ p_0.$$

The trickiest check of commutativity is that $\varepsilon^{B'} \circ q^B_0 = q^B \circ \varepsilon^B$. We write this mess out:

$$\begin{aligned} \varepsilon^{B'} \circ q^B_0 &= \varepsilon^{B'} \circ \iota'_0 \circ q^A_0 \circ k_0 + \varepsilon^{B'} \circ \iota'_0 \circ \alpha_0 \circ p_0 + \varepsilon^{B'} \circ j'_0 \circ q^C_0 \circ p_0 \\ &= f' \circ \varepsilon^{A'} \circ q^A_0 \circ k_0 + f' \circ \varepsilon^{A'} \circ \alpha_0 \circ p_0 + \varepsilon^{B'} \circ j'_0 \circ q^C_0 \circ p_0 \\ &= f' \circ q^A \circ \varepsilon^A \circ k_0 + f' \circ \beta_0 \circ p_0 + \varepsilon^{B'} \circ j'_0 \circ q^C_0 \circ p_0 \\ &= f' \circ q^A \circ \varepsilon^A \circ k_0 + (j'_0 \circ q^C \circ \varepsilon^C - \varepsilon^{B'} \circ j'_0 \circ q^C_0) \circ p_0 + \varepsilon^{B'} \circ j'_0 \circ q^C_0 \circ p_0 \\ &= f' \circ q^A \circ \varepsilon^A \circ k_0 + j'_0 \circ q^C \circ \varepsilon^C \circ p_0 \\ &= q^B \circ \varepsilon^B. \end{aligned}$$

The other q^B_i are defined similarly. For instance, one can see there exists a map $\beta_1 \colon P^C_1 \to P^{A'}_0$ such that

$$\iota_0 \circ \beta_1 = \iota_0 \circ \beta_0 \circ d^C_0 + j'_1 \circ q^C_0 \circ d^C_0 - d^{B'}_0 \circ j'_1 \circ q^C_1 \circ p_1,$$

and we set

$$q^B_1 = \iota'_1 \circ q^A_1 \circ k_1 + \iota'_1 \circ \alpha_1 \circ p_1 + j'_1 \circ q^C_1.$$

□

DEFINITION 12.7.10. A (homological) *universal δ-functor* is a δ-functor $F. = (F_i, \delta_i)$ with $F_i\colon \mathscr{C} \to \mathscr{D}$ such that if $G. = (G_i, \delta_i')$ is any other δ-functor with $G_i\colon \mathscr{C} \to \mathscr{D}$ for which there exists a natural transformation $\eta_0\colon G_0 \rightsquigarrow F_0$, then η_0 extends uniquely to a *morphism of δ-functors*, i.e., a sequence of natural transformations $\eta_i\colon G_i \rightsquigarrow F_i$ such that

$$\begin{tikzcd}
G_i(C) \arrow[r, "\delta_i'"] \arrow[d, "(\eta_i)_C"] & G_{i-1}(A) \arrow[d, "(\eta_{i-1})_A"] \\
F_i(C) \arrow[r, "\delta_i"] & F_{i-1}(A)
\end{tikzcd}$$

commutes for any short exact sequence in $\mathscr{C}$:

$$0 \to A \to B \to C \to 0.$$

(That is, we get a morphism of the associated long exact sequences.)

The δ-functor of left derived functors of F is universal, which will follow as a corollary of Theorem 12.7.14 below.

THEOREM 12.7.11. *The δ-functor (L_iF, δ_i) is universal.*

DEFINITION 12.7.12. Let $F\colon \mathscr{C} \to \mathscr{D}$ be a left exact functor between abelian categories. We say that an object Q in $\mathscr{C}$ is F-acyclic if $L_iF(Q) = 0$ for all $i \geq 1$.

Note that the $L_iF(A)$ for any $A \in \mathrm{Obj}(\mathscr{C})$ may be computed using resolutions by F-acyclic objects, as opposed to just projectives.

PROPOSITION 12.7.13. *Let $F\colon \mathscr{C} \to \mathscr{D}$ be a left exact functor between abelian categories, and let A be an object of $\mathscr{C}$. Suppose that $C. \to A$ is a resolution of A by F-acyclic objects. Then $L_iF(A) \cong H_i(F(C.))$ for each $i \geq 0$.*

PROOF. Note that we have an exact sequence

$$F(C_1) \xrightarrow{F(d_1^C)} F(C_0) \xrightarrow{F(\varepsilon^C)} F(A) \to 0,$$

so $F(A) \cong H_0(F(C.))$. Set $K_0 = \ker \varepsilon^C$. We then have an exact sequence

$$0 \to L_1F(A) \to F(K_0) \to F(C_0) \to F(A) \to 0,$$

which yields

$$L_1F(A) \cong \ker(\operatorname{coker}(F(C_2) \to F(C_1)) \to F(C_0)) \cong \frac{\ker F(d_1^C)}{\operatorname{im} F(d_2^C)} \cong H_1(F(C.)).$$

We also have isomorphisms $L_iF(A) \cong L_{i-1}F(K_0)$ for each $i \geq 2$.

For $i \geq 1$, set $K_i = \ker d_i^C \cong \operatorname{im} d_{i+1}^C$. The exact sequences

$$0 \to K_i \to C_i \to K_{i-1} \to 0,$$

then yield isomorphisms used in the following for $i \geq 2$:

$$L_iF(A) \cong L_{i-1}F(K_0) \cong \cdots \cong L_1F(K_{i-2}) \cong \ker(F(K_{i-1}) \to F(C_{i-1})) \cong \frac{\ker F(d_i^C)}{\operatorname{im} F(d_{i+1}^C)} \cong H_i(F(C.)).$$

□

More generally, we have the following characterization of universal δ-functors.

THEOREM 12.7.14. *Let $\mathscr{C}$ and $\mathscr{D}$ be abelian categories such that $\mathscr{C}$ has enough projectives. Suppose that (F_i, δ_i) form a δ-functor $F_i \colon \mathscr{C} \to \mathscr{D}$ and $F_i(P) = 0$ for every projective $P \in \mathrm{Obj}(\mathscr{C})$ and $i \geq 1$. Then (F_i, δ_i) is universal.*

PROOF. Suppose that (G_i, δ_i') is another δ-functor and that we have a natural transformation $G_0 \rightsquigarrow F_0$. Let $A \in \mathrm{Obj}(\mathscr{C})$ and let $\pi \colon P \to A$ be an epimorphism with P projective. Let $K = \ker \pi$. Let $i \geq 1$, and suppose that we have constructed a natural transformation $G_{i-1} \rightsquigarrow F_{i-1}$. Since $F_i(P) = 0$ is projective, we have a commutative diagram

$$\begin{tikzcd}
 & G_i(A) \arrow[r] \arrow[d, dashed] & G_{i-1}(K) \arrow[r] \arrow[d] & G_{i-1}(P) \arrow[d] \\
0 \arrow[r] & F_i(A) \arrow[r] & F_{i-1}(K) \arrow[r] & F_{i-1}(P).
\end{tikzcd}$$

The morphism $G_i(A) \to F_i(A)$ is the unique map which makes the diagram commute.

Now let $f \colon A \to B$ be a morphism in $\mathscr{C}$. We create a diagram as follows:

$$\begin{tikzcd}
0 \arrow[r] & K \arrow[r] \arrow[d] & P \arrow[r] \arrow[d] & A \arrow[r] \arrow[d, "f"] & 0 \\
0 \arrow[r] & K' \arrow[r] & P' \arrow[r] & B \arrow[r] & 0,
\end{tikzcd}$$

by taking P' to be projective, P to be any projective with an epimorphism to the pullback of the diagram $P' \to B \leftarrow A$, and K and K' to be the relevant kernels. We then have a diagram

$$\begin{tikzcd}
G_i(A) \arrow[rr] \arrow[dr] \arrow[dd] & & G_{i-1}(K) \arrow[dr] \arrow[dd] & \\
 & F_i(A) \arrow[rr, crossing over] & & F_{i-1}(K) \arrow[dd] \\
G_i(B) \arrow[rr] \arrow[dr] & & G_{i-1}(K') \arrow[dr] & \\
 & F_i(B) \arrow[rr] \arrow[from=uu, crossing over] & & F_{i-1}(K').
\end{tikzcd}$$

We need only see that the leftmost square commutes, but this follows easily from a diagram chase and the fact that the two horizontal maps on the frontmost square are monomorphisms.

Hence, we have constructed a sequence of natural transformations $G_i \rightsquigarrow F_i$. It remains only to see that these form a morphism of δ-functors. This being an inductive argument of the above sort, we leave it to the reader. □

As a corollary, we have a natural isomorphism of δ-functors between the left derived functors L_iF_0 of a right exact functor F_0 and any δ-functor (F_i, δ_i) with $F_i(P) = 0$ for P projective and $i \geq 1$.

We next wish to study right derived functors of left exact functors.

DEFINITION 12.7.15. An *injective resolution* of an object A of an abelian category is a cochain complex $I^\cdot$ of injective objects with $I^i = 0$ for $i < 0$ and a morphism $A \to I^0$ such that the resulting diagram

$$0 \to A \to I^0 \to I^1 \to I^2 \to \cdots$$

is exact.

DEFINITION 12.7.16. We say that an abelian category $\mathscr{C}$ has *enough* (or *sufficiently many*) injectives if for every $A \in \mathrm{Obj}(\mathscr{C})$, there exists an injective object $I \in \mathrm{Obj}(\mathscr{C})$ and a monomorphism $A \to I$.

REMARK 12.7.17. An abelian category has enough injectives if and only if every object of it has an injective resolution.

PROPOSITION 12.7.18. *The category R-**mod** has enough injectives.*

PROOF. First take the case that $R = \mathbb{Z}$. Let A be an abelian group, and write it as a quotient of a free abelian group

$$A = (\bigoplus_{j\in J} \mathbb{Z})/T$$

for some indexing set J and submodule T of $\bigoplus_{j\in J} \mathbb{Z}$. Then we may embed A in

$$I = (\bigoplus_{j\in J} \mathbb{Q})/T,$$

which is divisible as a quotient of a divisible group.

Next, let A be a left R-module. We have an injection of left R-modules,

$$\phi : A \to \mathrm{Hom}_{\mathbb{Z}}(R, A),$$

by $\phi(a)(r) = ra$. Now, embed A in a divisible group D, so that the resulting map

$$\mathrm{Hom}_{\mathbb{Z}}(R, A) \to \mathrm{Hom}_{\mathbb{Z}}(R, D)$$

is an injection. The proof that $\mathrm{Hom}_{\mathbb{Z}}(R, D)$ is an injective R-module is left to the reader. □

We also have the analogues of Propositions 12.6.5 and 12.6.6 for injective resolutions.

Suppose now that $F : \mathscr{C} \to \mathscr{D}$ is a left exact functor between abelian categories and that $\mathscr{C}$ has enough injectives. For each $i \geq 0$, we define additive functors $R^iF : \mathscr{C} \to \mathscr{D}$ by

$$R^iF(A) = H^i(F(I^\cdot)),$$

where $A \to I^\cdot$ is any injective resolution of $A \in \mathrm{Obj}(\mathscr{C})$ and, for $f : A \to B$ in $\mathscr{C}$, by

$$R^iF(f) : R^iF(A) \to R^iF(B)$$

to be the map on homology induced by any morphism of chain complexes $I^\cdot \to J^\cdot$ extending f, where $A \to I^\cdot$ and $B \to J^\cdot$ are injective resolutions. We have $R^0F = F$. The functors $R^\cdot F$ are called the *right-derived functors* of F.

DEFINITION 12.7.19. A *cohomological δ-functor* is a sequence of additive functors $F^i\colon \mathscr{C} \to \mathscr{D}$ for $i \in \mathbb{Z}$, together with, for every exact sequence

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

in $\mathscr{C}$, morphisms $\delta_i\colon F^i(C) \to F^{i+1}(A)$ fitting in a long exact sequence

$$\cdots \to F^i(A) \xrightarrow{F^i(f)} F^i(B) \xrightarrow{F^i(g)} F^i(C) \xrightarrow{\delta_i} F^{i+1}(A) \to \cdots$$

which are natural in the sense that if we have a morphism of short exact sequences in $\mathscr{C}$,

$$\begin{tikzcd}
0 \arrow[r] & A \arrow[r] \arrow[d] & B \arrow[r] \arrow[d] & C \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & A' \arrow[r] & B' \arrow[r] & C' \arrow[r] & 0,
\end{tikzcd}$$

then we obtain a morphism of long exact sequences in $\mathscr{D}$,

$$\begin{tikzcd}
\cdots \arrow[r] & F^i(A) \arrow[r] \arrow[d] & F^i(B) \arrow[r] \arrow[d] & F^i(C) \arrow[r] \arrow[d] & F^{i+1}(A) \arrow[r] \arrow[d] & \cdots \\
\cdots \arrow[r] & F^i(A') \arrow[r] & F^i(B') \arrow[r] & F^i(C') \arrow[r] & F^{i+1}(A') \arrow[r] & \cdots .
\end{tikzcd}$$

REMARK 12.7.20. A cohomological δ-functor (F^i, δ^i) is *universal* if there exists a unique extension of any natural transformation $F^0 \rightsquigarrow G^0$, where $(G^i, (\delta')^i)$ is another δ-functor, to a morphism of δ-functors.

THEOREM 12.7.21. *The functors $R^{\cdot}F$ form a cohomological universal δ-functor.*

The proof is dual to that of Theorems 12.7.9 and 12.7.11. We also have the following.

THEOREM 12.7.22. *Let $\mathscr{C}$ be an abelian category that has enough injectives. Then the cohomology functors $H^i\colon$ $\mathbf{Ch}^{\geq 0}(\mathscr{C}) \to \mathscr{C}$ for $i \geq 0$ on complexes in nonnegative degrees together with the connecting homomorphisms δ^i attached to a short exact sequence of complexes form a universal δ-functor.*

12.8. Tor and Ext

EXAMPLE 12.8.1. Take the abelian group $M = \mathbb{Z}/n\mathbb{Z}$. If we apply the functor $t_M\colon \mathbf{Ab} \to \mathbf{Ab}$ to the exact sequence of abelian groups $0 \to \mathbb{Z} \xrightarrow{n} \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z} \to 0$ for some $n \geq 2$, we obtain the right, but not left, exact sequence

$$\mathbb{Z}/n\mathbb{Z} \xrightarrow{0} \mathbb{Z}/n\mathbb{Z} \xrightarrow{\sim} \mathbb{Z}/n\mathbb{Z} \to 0.$$

If we apply $h_M\colon \mathbf{Ab} \to \mathbf{Ab}$ to the same exact sequence, we obtain the left, but not right, exact sequence

$$0 \to 0 \to 0 \to \mathbb{Z}/n\mathbb{Z},$$

noting that $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/n\mathbb{Z}, \mathbb{Z}) = 0$.

DEFINITION 12.8.2. A right R-module N is *R-flat*, or just *flat*, if the tensor product functor $t_N\colon R\text{-}\mathbf{mod}\to\mathbf{Ab}$ is exact.

REMARKS 12.8.3.

a. An S-R-bimodule N is flat as a right R-module if and only if the functor $t_N\colon R\text{-}\mathbf{mod}\to S\text{-}\mathbf{mod}$ is exact.

b. A left R-module M is defined to be flat if it is flat as a right R^{op}-module (which is equivalent to the right tensor product functor with M being exact on $R^{\mathrm{op}}\text{-}\mathbf{mod}$).

PROPOSITION 12.8.4. *Projective right R-modules are R-flat.*

PROOF. Let P be a projective R-module, and let Q be a complement in a free R-module F on a basis X. Let $f\colon A\to B$ be an injection of R-modules. We have a commutative diagram

$$\begin{tikzcd}
P\otimes_R A \arrow[r, "\mathrm{id}_P\otimes f"] \arrow[d, hook] & P\otimes_R B \arrow[d, hook] \\
F\otimes_R A \arrow[r, "\mathrm{id}_F\otimes f"] \arrow[d, "\wr"] & F\otimes_R B \arrow[d, "\wr"] \\
\bigoplus_{x\in X} A \arrow[r, hook, "(f)_{x\in X}"] & \bigoplus_{x\in X} B,
\end{tikzcd}$$

the vertical isomorphisms following from the commutativity of direct sums and tensor products. Since f is injective, so is the lowermost vertical map. Since P is a direct sum of F, the map

$$P\otimes_R A\to(P\otimes_R A)\oplus(Q\otimes_R A)\xrightarrow{\sim}F\otimes_R A$$

is injective, and similarly with A replaced by B. Thus, commutativity of the diagram yields the injectivity $\mathrm{id}_P\otimes f$. Since left tensor product with P preserves injective homomorphisms and right exact sequences, it preserves short exact sequences and is therefore exact. □

For modules over a principal ideal domain, we can characterize flat modules as follows.

PROPOSITION 12.8.5. *Let R be a PID. An R-module M is flat if and only if M is R-torsion-free.*

PROOF. Let M be a flat R-module. Let $r\in R$ be a nonzero element, and let $\phi_r\colon R\to R$ be the injective map $\phi_r(x)=rx$ for $x\in R$. The tensor product map $\mathrm{id}_M\otimes_R\phi_r$ is injective as A is R-flat. Under the identification $M\otimes_R R\cong M$ of Corollary 9.3.24, determined by $m\otimes r\mapsto rm$, the map $\mathrm{id}_M\otimes_R\phi_r$ becomes identified with the map $\psi_r\colon M\to M$ that is left multiplication by r. Since ψ_r is then injective for every nonzero r, we see that M has no nonzero R-torsion.

Next, let M be R-torsion free. It is the union (which is also the direct limit) of its finitely generated, necessarily torsion-free R-submodules. We omit here a check of the fact that direct limits and tensor products commute. Given this, we may assume that M is finitely generated, in which case it follows from Proposition 12.8.5 that M is free, hence projective, and hence flat. □

REMARK 12.8.6. It follows from Corollary 9.9.3 and Proposition 12.8.5 that finitely generated flat modules over a PID R are R-free.

DEFINITION 12.8.7. Let R and S be rings, and let A be an S-R-bimodule. For $i \geq 0$, the ith Tor-functor

$$\mathrm{Tor}_i^R(A,\cdot)\colon R\text{-}\mathbf{mod} \to S\text{-}\mathbf{mod}$$

is the ith left derived functor of t_A.

REMARK 12.8.8. If R is a commutative ring, then an R-module A provides functors

$$\mathrm{Tor}_i^R(A,\cdot)\colon R\text{-}\mathbf{mod} \to R\text{-}\mathbf{mod}$$

since R-modules are automatically R-R-bimodules.

REMARK 12.8.9. As $\mathrm{Tor}_i^R(A,B) = H_i(A \otimes_R Q.)$ for any projective resolution $Q.$ of B by R-modules, the composition of the functor

$$\mathrm{Tor}_i^R(A,\cdot)\colon R\text{-}\mathbf{mod} \to S\text{-}\mathbf{mod}$$

with the forgetful functor $F\colon R\text{-}\mathbf{mod} \to \mathbf{Ab}$ agrees with the functor

$$\mathrm{Tor}_i^R(F(A),\cdot)\colon R\text{-}\mathbf{mod} \to \mathbf{Ab},$$

hence the omission of the notation for S in the definition of $\mathrm{Tor}_i^R(A,\cdot)$.

EXAMPLE 12.8.10. In $\mathbf{Ab}$, consider the projective resolution

$$0 \to \mathbb{Z} \xrightarrow{n} \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z} \to 0$$

of B. Computing the homology of $0 \to A \xrightarrow{n} A \to 0$, we obtain

$$\mathrm{Tor}_i^{\mathbb{Z}}(A,\mathbb{Z}/n\mathbb{Z}) \cong \begin{cases} A/nA & \text{if } i = 0 \\ A[n] = \{a \in A \mid na = 0\} & \text{if } i = 1 \\ 0 & \text{if } i \geq 2. \end{cases}$$

LEMMA 12.8.11. *Let R be a ring. The following conditions on a right R-module A are equivalent:*

i. A is flat,

ii. $\mathrm{Tor}_1^R(A,\cdot) = 0$*,*

iii. $\mathrm{Tor}_i^R(A,\cdot) = 0$ *for all $i \geq 1$.*

PROOF. Clearly, (iii) implies (ii). If

$$0 \to B_1 \to B_2 \to B_3 \to 0$$

is an exact sequence of right R-modules, then we have a long exact sequence for any R-module that ends with

$$\mathrm{Tor}_1^R(A,B_3) \to A \otimes_R B_1 \to A \otimes_R B_2 \to A \otimes_R B_3 \to 0,$$

from which it is clear that (ii) implies (i).

Finally, if (i) holds and $Q.$ is a projective resolution of B in R-$\mathbf{mod}$, then the complex

$$\cdots \to A \otimes_R Q_1 \to A \otimes_R Q_0 \to A \otimes_R B \to 0$$

is exact by the flatness of A. It follows that

$$\operatorname{Tor}_i^R(A,B) = H_i(A \otimes_R Q.) = 0$$

for all $i \geq 1$. □

PROPOSITION 12.8.12. *Let A be a right R-module and B a left R-module. Let $P. \to A$ be a resolution of A by projective right R-modules. Then*

$$\operatorname{Tor}_i^R(A,B) \cong H_i(P. \otimes_R B)$$

for all $i \geq 0$. In particular, the functors $\operatorname{Tor}_i^R(\,\cdot\,,B)$ are the left derived functors of R-tensor product with B.

PROOF. We sketch a proof. Form projective resolutions $P. \to A$ and $Q. \to B$. We then have a double complex $P. \otimes_R Q.$, and we can consider homology of the total complex

$$\operatorname{Tot}(P. \otimes_R Q.)_k = \bigoplus_{i+j=k} P_i \otimes_R Q_j,$$

where the boundary maps from each term $P_i \otimes_R Q_j$ are given by the sums

$$d_i^A \otimes \mathrm{id}_{Q_j} + (-1)^i \mathrm{id}_{P_i} \otimes d_j^B.$$

We claim that the homology of this chain complex is isomorphic to the homology of the complexes $P. \otimes_R B$ and $A \otimes_R Q.$, from which the lemma follows.

We have maps of complexes

$$\operatorname{Tot}(P. \otimes_R Q.) \to P. \otimes_R B \tag{12.8.1}$$

and

$$\operatorname{Tot}(P. \otimes_R Q.) \to A \otimes_R Q. \tag{12.8.2}$$

induced by augmentation morphisms (up to sign, and zero maps otherwise). The double complex $P. \otimes_R Q. \to P. \otimes_R B$ (i.e., with $P_i \otimes_R B$ in the $(i,-1)$-position) has exact columns, since each projective module is flat. One can show that this implies that the total complex of this cpomplex is exact. This says precisely that the map in (12.8.1) induces an isomorphism on homology. Similarly, so does the map in (12.8.2). □

We have the following almost immediate corollary, since left and right tensor product with a module over a commutative ring are naturally isomorphic functors.

COROLLARY 12.8.13. *Let R be commutative. We have $\operatorname{Tor}_i^R(A,B) \cong \operatorname{Tor}_i^R(B,A)$ for all R-modules A, B and $i \geq 0$.*

We now give an alternate proof of Proposition 12.8.12.

PROOF. Let $Q. \to B$ be a projective resolution of A by right R-modules. Suppose that

$$0 \to A_1 \to A_2 \to A_3 \to 0$$

is an exact sequence. Then

$$0 \to A_1 \otimes_R Q. \to A_2 \otimes_R Q. \to A_3 \otimes_R Q. \to 0$$

is exact. This yields a long exact sequence in homology of the form

$$\cdots \to \mathrm{Tor}_i^R(A_1,B) \to \mathrm{Tor}_i^R(A_2,B) \to \mathrm{Tor}_i^R(A_3,B) \to \mathrm{Tor}_{i-1}^R(A_1,B) \to \cdots,$$

so the functors $\mathrm{Tor}_i^R(\,\cdot\,,B)$ do in fact form a δ-functor. Futhermore, since any projective right R-module P is flat, we have that $\mathrm{Tor}_i^R(P,B) = 0$ for all $i \geq 1$. By Theorem 12.7.14, it follows that the $\mathrm{Tor}_i^R(\,\cdot\,,B)$ are a universal δ-functor extending t_B. The proposition therefore follows by Theorem 12.7.11. □

REMARK 12.8.14. It follows from Proposition 12.8.12 and Proposition 12.7.13 that the $\mathrm{Tor}_i^R(A,B)$ can be computed via a flat resolution of either A or B.

The following explains something more of the name "Tor".

LEMMA 12.8.15. *The functor* $\mathrm{Tor}_1^{\mathbb{Z}}(A,\,\cdot\,) = 0$ *if and only if A is torsion-free.*

PROOF. We prove this for finitely generated abelian groups. (The general result then follows from the fact that left derived functors commute with colimits.) By Proposition 12.8.13, we may compute $\mathrm{Tor}_1^{\mathbb{Z}}(A,B)$ by finding a projective resolution of A. Say

$$A \cong \mathbb{Z}^m \oplus \mathbb{Z}/n_1\mathbb{Z} \oplus \cdots \mathbb{Z}/n_r\mathbb{Z}$$

with $r \geq 0$ and the $n_i \geq 2$. Then we have a projective resolution of the form

$$0 \to \mathbb{Z}^{m+r} \xrightarrow{(1,\cdots,1,n_1,\cdots,n_r)} \mathbb{Z}^{m+r} \to A \to 0.$$

Tensoring with B and computing H_1, we obtain $B[n_1] \oplus \cdots \oplus B[n_r]$. This will always be trivial if and only if $r = 0$. □

By Lemma 12.8.15, a $\mathbb{Z}$-module is flat if and only if it is torsion-free. This is seen to hold in the same manner with $\mathbb{Z}$ replaced by any PID. Note that this does not hold for all commutative rings.

EXAMPLE 12.8.16. Consider $R = \mathbb{Q}[x,y]$. Then the exact sequence

$$0 \to R \xrightarrow{(y,-x)} R^2 \xrightarrow{(a,b)\mapsto ax+by} R \to \mathbb{Q} \to 0$$

is a free resolution of $\mathbb{Q}$. Let J be the ideal (x,y) of R, so $\mathbb{Q} \cong R/J$. Then we have isomorphisms

$$\mathrm{Tor}_1^R(J,\mathbb{Q}) \cong \mathrm{Tor}_2^R(\mathbb{Q},\mathbb{Q}) \cong \ker(\mathbb{Q} \xrightarrow{0} \mathbb{Q}^2) = \mathbb{Q}.$$

Thus J is not flat as an R-module, even though it is torsion-free.

Here is another class of examples.

LEMMA 12.8.17. *Let S be a subset of R that is multiplicatively closed. Then the localization $S^{-1}R$ is a flat R-module.*

PROOF. Recall that we have natural isomorphisms $S^{-1}A \cong S^{-1}R \otimes_R A$ for R-modules A. Suppose that $f\colon A \to B$ is an injection of R-modules. Then we obtain an induced R-module homomorphism $\tilde{f}\colon S^{-1}A \to S^{-1}B$, which we must show is an injection. Suppose $\tilde{f}(s^{-1}a) = 0$. Then

$$0 = s\tilde{f}(s^{-1}a) = \tilde{f}(a) = f(a),$$

so $a = 0$. □

DEFINITION 12.8.18. Let R and S be rings, and let A be an R-S-bimodule. For $i \geq 0$, the ith Ext-functor

$$\operatorname{Ext}^i_R(A, \cdot)\colon R\textbf{-mod} \to S\textbf{-mod}$$

is the ith right derived functors of h_A.

EXAMPLE 12.8.19. For $R = \mathbb{Z}$, we may consider the injective resolution

$$0 \to \mathbb{Z}/n\mathbb{Z} \to \mathbb{Q}/\mathbb{Z} \xrightarrow{n} \mathbb{Q}/\mathbb{Z} \to 0$$

of $\mathbb{Z}/n\mathbb{Z}$. For any abelian group B, we write $B^\vee = \operatorname{Hom}(B, \mathbb{Q}/\mathbb{Z})$. We must compute the cohomology of $A^\vee \xrightarrow{n} A^\vee$. This yields

$$\operatorname{Ext}^i_{\mathbb{Z}}(A, \mathbb{Z}/n\mathbb{Z}) \cong \begin{cases} A^\vee[n] & \text{if } i = 0 \\ A^\vee/nA^\vee & \text{if } i = 1 \\ 0 & \text{if } i = 2. \end{cases}$$

One has that $\operatorname{Ext}^i_R(P,B) = 0$ for all B and all $i \geq 1$ if P is a projective module, as follows from the exactness of $\operatorname{Hom}_R(P, \cdot)$. We have the analogous result to Proposition 12.8.13 for Ext-groups, which says that such groups may be computed using projective resolutions.

PROPOSITION 12.8.20. *We have* $\operatorname{Ext}^i_R(A,B) \cong H^i(\operatorname{Hom}_R(P_\cdot, B))$*, where* $P_\cdot \to A$ *is any projective resolution of* A*.*

We end with a characterization of Ext^1_R in terms of extensions.

DEFINITION 12.8.21. An *extension* of an R-module A by an R-module B is an exact sequence $0 \to B \to E \to A \to 0$, where E is an R-module. Two extensions of A by B are called equivalent if there is an isomorphism of exact sequences between them that is the identity on A and B.

Note that all split extensions (i.e., those with split exact sequences) are split.

EXAMPLE 12.8.22. There are p equivalence classes of extensions of $\mathbb{Z}/p\mathbb{Z}$ by $\mathbb{Z}/p\mathbb{Z}$ as $\mathbb{Z}$-modules:

$$0 \to \mathbb{Z}/p\mathbb{Z} \xrightarrow{pi} \mathbb{Z}/p^2\mathbb{Z} \xrightarrow{\bmod p} \mathbb{Z}/p\mathbb{Z} \to 0$$

with $1 \leq i \leq p-1$, and

$$0 \to \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z} \to 0.$$

THEOREM 12.8.23. *There is a one-to-one correspondence between equivalence classes of extensions of* A *by* B *and* $\operatorname{Ext}^1_R(A,B)$.

PROOF. Suppose that $\mathscr{E}$ is an equivalence class of extensions of A by B, representative by an exact sequence

$$0 \to B \to E \to A \to 0. \tag{12.8.3}$$

We then have an exact sequence

$$\operatorname{Hom}_R(E,B) \to \operatorname{Hom}_R(B,B) \xrightarrow{\partial_{\mathscr{E}}} \operatorname{Ext}^1_R(A,B),$$

and we set $\Phi(\mathscr{E}) = \partial_{\mathscr{E}}(\mathrm{id}_B)$. This is clearly independent of the choice of representative.

Conversely, suppose $u \in \operatorname{Ext}^1_R(A,B)$. Fix an exact sequence

$$0 \to K \xrightarrow{\iota} P \to A \to 0$$

with P projective. We then have an exact sequence

$$\operatorname{Hom}_R(P,B) \to \operatorname{Hom}_R(K,B) \xrightarrow{\partial} \operatorname{Ext}^1_R(A,B) \to 0.$$

Let $t \in \operatorname{Hom}_R(K,B)$ with $\partial(t) = u$. Let E be the pushout

$$E = P \amalg_K B = P \oplus B / \{(\iota(k), t(k)) \mid k \in K\}.$$

We have a commutative diagram

(12.8.4)
$$\begin{tikzcd}
0 \arrow[r] & K \arrow[r] \arrow[d, "t"] & P \arrow[r] \arrow[d] & A \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & B \arrow[r] & E \arrow[r] & A \arrow[r] & 0.
\end{tikzcd}$$

Here, the map $E \to A$ is defined by universality of the pushout (via the map $P \to A$ and the zero map $B \to A$). We define $\Psi(u)$ to be the equivalence class $\mathscr{E}'$ of the extension given by the lower row. Though it is not immediately clear that this is independent of the choice of t with $\partial(t) = u$, this follows if we can show that Ψ and Φ as constructed are mutually inverse.

To see that $\Phi(\Psi(u)) = u$, set $\mathscr{E} = \Psi(u)$, again choosing any t with $\partial(t) = u$. The diagram

(12.8.5)
$$\begin{tikzcd}
\operatorname{Hom}_R(B,B) \arrow[r, "\partial_{\mathscr{E}}"] \arrow[d, "h^B(t)"] & \operatorname{Ext}^1_R(A,B) \arrow[d, equal] \\
\operatorname{Hom}_R(K,B) \arrow[r, "\partial"] & \operatorname{Ext}^1_R(A,B),
\end{tikzcd}$$

commutes. Hence, we have

$$\Phi(\Psi(u)) = \Phi(\mathscr{E}) = \partial_{\mathscr{E}}(\mathrm{id}_B) = \partial(t) = u,$$

as desired.

On the other hand, suppose given $\mathscr{E}$ with exact sequence (12.8.3). By projectivity of P, the map $P \to A$ lifts to a map $P \to E$. Hence, we have a diagram as in (12.8.4). Furthermore, the map t in the diagram (12.8.4) satisfies $\partial(t) = \partial_{\mathscr{E}}(\mathrm{id}_B)$ by the commutativity of (12.8.5). Now, there exists a map $P \amalg_K B \to E$ by universality of the pushout, and it is the identity on A and B, hence an isomorphism by the 5-lemma. It follows by construction that

$$\Psi(\Phi(\mathscr{E})) = \Psi(\partial_{\mathscr{E}}(\mathrm{id}_B)) = \mathscr{E}.$$

□

12.9. Group cohomology

In this section, we let G denote a group.

DEFINITION 12.9.1. The *augmentation map* $\varepsilon \colon \mathbb{Z}[G] \to \mathbb{Z}$ is the unique ring homomorphism with $\varepsilon(g) = 1$ for all $g \in G$.

DEFINITION 12.9.2. The *augmentation ideal I_G* of $\mathbb{Z}[G]$ is the kernel of the augmentation map.

LEMMA 12.9.3. *The augmentation ideal I_G is generated by* $\{g-1 \mid g \in G\}$.

PROOF. We have

$$I_G = \left\{ \sum_{g\in G} a_g g \in \mathbb{Z}[G] \mid \sum_{g\in G} a_g = 0 \right\}.$$

For $\alpha = \sum_{g\in G} a_g g \in I_G$, we have

$$\alpha = \alpha - \sum_{g\in G} a_g = \sum_{g\in G} a_g(g-1).$$

□

DEFINITION 12.9.4. Let A be an $\mathbb{Z}[G]$-module.

a. The *G-invariant group* of A is the $\mathbb{Z}$-module

$$A^G = \{a \in A \mid ga = a \text{ for all } g \in G\},$$

the maximal $\mathbb{Z}[G]$-submodule of A on which all elements of G act trivially.

b. The *G-coinvariant group* of A is the $\mathbb{Z}$-module $A_G = A/I_GA$, the maximal $\mathbb{Z}[G]$-quotient of A on which all elements of G act trivially.

EXAMPLES 12.9.5.

a. If we view $\mathbb{Z}$ as a $\mathbb{Z}[G]$-trivial module, we have $\mathbb{Z}^G = \mathbb{Z}$ and $\mathbb{Z}_G \cong \mathbb{Z}$.

b. We have $\mathbb{Z}[G]_G \cong \mathbb{Z}$ via the augmentation map, and

$$\mathbb{Z}[G]^G = \begin{cases} \mathbb{Z}\cdot N_G & \text{if G is finite} \\ 0 & \text{otherwise,} \end{cases}$$

where $N_G = \sum_{g\in G} g$ is the *norm element* in a finite group G. The computation of the invariant group follows from the fact that the action of G on itself by left multiplication is transitive, so for an element of $\mathbb{Z}[G]$ to be G-fixed, its coefficients must all be equal.

c. Let K/F be a finite Galois extension of fields, and let $G = \mathrm{Gal}(K/F)$. Then $K^G = F$ and $(K^\times)^G = F^\times$.

EXAMPLE 12.9.6. For $n \geq 2$, let S_n act on $\mathscr{A} = \mathbb{Z}[x_1, x_2, \dots, x_n]$ by

$$\sigma \cdot p(x_1, x_2, \dots, x_n) = p(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)})$$

for $\sigma \in S_n$ and $p \in \mathscr{A}$. This action is $\mathbb{Z}$-bilinear so it gives $\mathscr{A}$ the structure of a left $\mathscr{A}[S_n]$-module. Then $\mathscr{A}^{S_n}$ is the $\mathbb{Z}$-module of symmetric polynomials in $\mathscr{A}$, which is the $\mathbb{Z}$-module generated by the elementary symmetric polynomials (see Definition 6.13.4). On the other hand, $\mathscr{A}_{S_n} \cong \mathbb{Z}[x]$, with the isomorphism induced by the $\mathbb{Z}$-linear map $\mathscr{A} \to \mathbb{Z}[x]$ taking each x_i to x.

REMARK 12.9.7. We have a left exact invariant functor $A \mapsto A^G$ as a functor $\mathbb{Z}[G]$-**mod** $\to$ **Ab**, with the map on homomorphisms being the restriction to invariant subgroups. This functor is naturally isomorphic to the functor $h_{\mathbb{Z}}$, where $\mathbb{Z}$ is viewed as the trivial $\mathbb{Z}[G]$-module. In particular

$$\eta_A \colon \operatorname{Hom}_{\mathbb{Z}[G]}(\mathbb{Z},A) \to A^G, \quad \eta_A(\phi) = \phi(1)$$

for $\phi \in \operatorname{Hom}_{\mathbb{Z}[G]}(\mathbb{Z},A)$ is a natural isomorphism. Thus, the invariant factor is left exact.

Similarly, $A \mapsto A_G$ defines a right exact coinvariant functor which is isomorphic to the functor $t_{\mathbb{Z}}$, in that we have natural isomorphisms

$$\mathbb{Z} \otimes_{\mathbb{Z}[G]} A \xrightarrow{\sim} A_G, \qquad 1 \otimes a \mapsto a + I_G A.$$

In particular, the coinvariant functor is right exact.

DEFINITION 12.9.8.

a. The cohomology $H^*(G,\cdot)$ of G is the δ-functor given by the right derived functors of the G-invariant functor. The *ith cohomology group of G with coefficients in a $\mathbb{Z}[G]$-module A is $H^i(G,A)$.*

b. The homology $H_*(G,\cdot)$ of G is the δ-functor given by the left derived functor of the G-coinvariant functor. The *ith homology group of G with coefficients in a $\mathbb{Z}[G]$-module A is $H_i(G,A)$.*

REMARK 12.9.9. By definition, we have natural isomorphisms

$$H^i(G,A) \cong \operatorname{Ext}^i_{\mathbb{Z}[G]}(\mathbb{Z},A) \quad \text{and} \quad H_i(G,A) \cong \operatorname{Tor}_i^{\mathbb{Z}[G]}(\mathbb{Z},A)$$

for $i \geq 0$ and $\mathbb{Z}[G]$-modules A.

Let us give a more explicit description of group cohomology.

DEFINITION 12.9.10. The *bar resolution* of $\mathbb{Z}$ as a $\mathbb{Z}[G]$-module is the complex $C.$ with $C_i = \mathbb{Z}[G^{i+1}]$ for $i \geq 0$, differentials $d_i \colon C_i \to C_{i-1}$ given on $(g_0,\dots,g_i) \in G^{i+1}$ by

$$d_i((g_0,\dots,g_i)) = \sum_{j=0}^{i} (-1)^j (g_0,\dots,g_{j-1},g_{j+1},\dots,g_i)$$

and augmentation $\varepsilon \colon C_0 \to \mathbb{Z}$ the augmentation map.

REMARK 12.9.11. As follows from Remark 12.9.9, the group $H^i(G,A)$ is the ith cohomology group of the complex

$$0 \to \operatorname{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G],A) \xrightarrow{D^0} \operatorname{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G^2],A) \to \cdots$$
$$\to \operatorname{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G^i],A) \xrightarrow{D^{i-1}} \operatorname{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G^{i+1}],A) \to \cdots$$

with $\operatorname{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G],A) \cong A$ in degree 0. Similarly, $H_i(G,A)$ is the ith homology group of the complex

$$\cdots \to \mathbb{Z}[G^{i+1}] \otimes_{\mathbb{Z}[G]} A \to \cdots \to \mathbb{Z}[G^2] \otimes_{\mathbb{Z}[G]} A \to \mathbb{Z}[G] \otimes_{\mathbb{Z}[G]} A \to 0,$$

with $\mathbb{Z}[G] \otimes_{\mathbb{Z}[G]} A \cong A$ in degree 0.

There is another complex which computes the cohomology of G, that of the inhomogeneous G-cocycles, which has a more complicated differential but is more amenable to computation.

DEFINITION 12.9.12. Let A be a G-module, and let $i \geq 0$.

a. The group of *i-cochains* of G with coefficients in A is the set of functions from G^i to A:

$$C^i(G,A) = \{f \colon G^i \to A\}.$$

b. The ith *differential* $d^i = d^i_A \colon C^i(G,A) \to C^{i+1}(G,A)$ is the map

$$d^i(f)(g_0,g_1,\ldots,g_i) = g_0 \cdot f(g_1,\ldots g_i) \\ + \sum_{j=1}^{i} (-1)^j f(g_0,\ldots,g_{j-2},g_{j-1}g_j,g_{j+1},\ldots,g_i) + (-1)^{i+1} f(g_0,\ldots,g_{i-1}).$$

We remark that $C^0(G,A)$ is taken simply to be A, as G^0 is a singleton set. The proof of the following, which tells us that $C^{\cdot}(G,A)$ is a cochain complex, is left to the reader.

LEMMA 12.9.13. *For any $i \geq 0$, one has $d^{i+1} \circ d^i = 0$.*

We consider the cohomology groups of $C^{\cdot}(G,A)$.

DEFINITION 12.9.14. Let $i \geq 0$.

a. We set $Z^i(G,A) = \ker d^i$, the group of *i-cocycles* of G with coefficients in A.

b. We set $B^0(G,A) = 0$ and $B^i(G,A) = \operatorname{im} d^{i-1}$ for $i \geq 1$. We refer to $B^i(G,A)$ as the group of *i-coboundaries* of G with coefficients in A.

THEOREM 12.9.15. *The maps*

$$\psi^i \colon \operatorname{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G^{i+1}],A) \to C^i(G,A)$$

defined by

$$\psi^i(\varphi)(g_1,\ldots,g_i) = \varphi(1,g_1,g_1g_2,\ldots,g_1g_2\cdots g_i)$$

are isomorphisms for all $i \geq 0$. This provides isomorphisms of complexes in the sense that $\psi^{i+1} \circ D^i = d^i \circ \psi^i$ for all $i \geq 0$. Moreover, these isomorphisms are natural in the G-module A.

PROOF. If $\psi^i(\varphi) = 0$, then

$$\varphi(1,g_1,g_1g_2,\ldots,g_1g_2\cdots g_i) = 0$$

for all $g_1,\ldots,g_i \in G$. Let $h_0,\ldots,h_i \in G$, and define $g_j = h_{j-1}^{-1}h_j$ for all $1 \leq j \leq i$. We then have

$$\varphi(h_0,h_1,\ldots,h_i) = h_0\varphi(1,h_0^{-1}h_1,\ldots,h_0^{-1}h_i) = h_0\varphi(1,g_1,\ldots,g_1\cdots g_i) = 0.$$

Therefore, ψ^i is injective. On the other hand, if $f \in C^i(G,A)$, then defining

$$\varphi(h_0,h_1,\ldots,h_i) = h_0 f(h_0^{-1}h_1,\ldots,h_{i-1}^{-1}h_i),$$

we have

$$\varphi(gh_0,gh_1,\ldots,gh_i) = gh_0 f((gh_0)^{-1}gh_1,\ldots,(gh_{i-1})^{-1}gh_i) = g\varphi(h_0,h_1,\ldots,h_i)$$

and $\psi^i(\varphi) = f$. Therefore, ψ^i is an isomorphism of groups.

That $\psi^{\cdot}$ forms a map of complexes is shown in the following computation:

$$\begin{aligned}\psi^{i+1}(D^i(\varphi))(g_1,\dots,g_{i+1}) &= D^i(\varphi)(1,g_1,\dots,g_1\cdots g_{i+1})\\ &= \varphi\circ d_{i+1}(1,g_1,\dots,g_1\cdots g_{i+1})\\ &= \sum_{j=0}^{i+1}(-1)^j\varphi(1,g_1,\dots,g_1\cdots g_{j-2},g_1\cdots g_j,\dots,g_1\cdots g_{i+1}).\end{aligned}$$

The latter term equals

$$g_1\psi^i(\varphi)(g_2,\dots,g_{i+1})+\sum_{j=1}^{i}(-1)^j\psi^i(\varphi)(g_1,\dots,g_{j-2},g_{j-1}g_j,g_{j+1},\dots,g_{i+1}) \\ +(-1)^{i+1}\psi^i(\varphi)(g_1,\dots,g_i),$$

which is $d^i(\psi^i(\varphi))$.

Finally, suppose that $\alpha\colon A\to B$ is a G-module homomorphism. We then have

$$\alpha\circ\psi^i(\varphi)(g_1,\dots,g_i)=\alpha\circ\varphi(1,g_1,\dots,g_1\cdots g_i)=\psi^i(\alpha\circ\varphi)(g_1,\dots,g_i),$$

hence the desired naturality. □

COROLLARY 12.9.16. *The ith cohomology group of the complex* $(\mathrm{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G^{i+1}],A),D_A^i)$ *is naturally isomorphic to* $H^i(G,A)$.

COROLLARY 12.9.17. *The ith cohomology group of G with coefficients in A is*

$$H^i(G,A)=Z^i(G,A)/B^i(G,A).$$

The cohomology groups measure how far the cochain complex $C^{\cdot}(G,A)$ is from being exact. We give some examples of cohomology groups in low degree.

LEMMA 12.9.18. *We have*

$$Z^1(G,A)=\{f\colon G\to A\mid f(gh)=gf(h)+f(g)\,\textit{for all}\,g,h\in G\}$$

and $B^1(G,A)$ *is the subgroup of* $f\colon G\to A$ *for which there exists* $a\in A$ *such that* $f(g)=ga-a$ *for all* $g\in G$. *In particular, if A is a* $\mathbb{Z}[G]$*-module with trivial G-action, then* $H^1(G,A)=\mathrm{Hom}(G,A)$.

PROOF. Let $a\in A$. Then $d^0(a)(g)=ga-a$ for $g\in G$, so $\ker d^0=A^G$. That proves part a, and part b is simply a rewriting of the definitions. Part c follows immediately, as the definition of $Z^1(G,A)$ reduces to $\mathrm{Hom}(G,A)$, and $B^1(G,A)$ is clearly (0), in this case. □

We remark that, as A is abelian, we have $\mathrm{Hom}(G,A)=\mathrm{Hom}(G^{\mathrm{ab}},A)$, where G^{ab} is the maximal abelian quotient of G (i.e., its abelianization).

We turn briefly to an interesting use for second cohomology groups.

DEFINITION 12.9.19. A *group extension* of G by a G-module A is a short exact sequence of groups

$$0\to A\xrightarrow{\iota}\mathscr{E}\xrightarrow{\pi}G\to 1$$

such that, choosing any section $s\colon G \to \mathscr{E}$ of π, one has

$$s(g)as(g)^{-1} = g \cdot a$$

for all $g \in G$, $a \in A$. Two such extensions $\mathscr{E} \to \mathscr{E}'$ are said to be equivalent if there is an isomorphism $\theta\colon \mathscr{E} \xrightarrow{\sim} \mathscr{E}'$ fitting into a commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & A \arrow[r] \arrow[d, equal] & \mathscr{E} \arrow[r] \arrow[d, "\theta"] & G \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & A \arrow[r] & \mathscr{E}' \arrow[r] & G \arrow[r] & 0,
\end{tikzcd}$$

We denote the set of equivalence classes of such extensions by $\mathscr{E}(G,A)$.

DEFINITION 12.9.20. A *factor set* of a group G valued in a $\mathbb{Z}[G]$-module A is a 2-cocycle $f\colon G^2 \to A$ satisfying $f(1,g) = f(g,1) = 0$ for all $g \in G$.

LEMMA 12.9.21. *Every* 2*-cocycle of a group is cohomologous to, i.e., has the same cohomology class as, a factor set.*

PROOF. The condition that $F\colon G^2 \to A$ is a 2-cocycle is that

$$gF(h,k) + F(g,hk) = F(gh,k) + F(g,h)$$

for all $g,h,k \in G$. In particular, taking $g = h = e$, we have $F(e,k) = F(e,e)$ and taking $h = k = e$, we have $gF(e,e) = fFg,e)$. Note that for a 1-cochain c, we have

$$dc(g,h) = gc(h) - c(gh) + c(g).$$

In particular, if we set $c(g) = c$ for all $g \in G$ some fixed $c \in A$, then $dc(g,h) = gc$ for all $g \in G$, so if we take $c = F(e,e)$ and replace F by $f = F - dc$, then $f(e,k) = 0$ and $f(g,e) = 0$ for all $g,k \in G$. □

THEOREM 12.9.22. *The group $H^2(G,A)$ is in canonical bijection with $\mathscr{E}(G,A)$ via the map induced by that taking a factor set $f\colon G^2 \to A$ to the extension $\mathscr{E}_f = A \times G$ with multiplication given by*

$$(a,g) \cdot (b,h) = (a + gb + f(g,h), gh)$$

This identification takes the identity to the semi-direct product $A \ltimes G$ determined by the action of G on A.

PROOF. We check that $\mathscr{E}_f$ so defined is a group. That it has identity $(0,e)$ is clear from the definition. Associativity is as follows:

$$\begin{aligned}(a + gb + f(g,h), gh) \cdot (c,k) &= (a + gb + f(g,h) + ghc + f(gh,k), ghk) \\ &= (a + g(b + hc) + gf(h,k) + f(g,hk), ghk) = (a,g) \cdot (b + hc + f(h,k), hk).\end{aligned}$$

The inverse of (a,g) clearly has the form (b,g^{-1}) for some $b \in A$, and we then must have $a + gb + f(g,g^{-1}) = 0$, so $b = -g^{-1}a - g^{-1}f(g,g^{-1})$, and the inverse exists. That $\mathscr{E}_f$ is a group extension of G by A is now nearly immediate. Note also that $\mathscr{E}_f$ is split if $f = 0$,]since in that case $(0,g)(0,h) = (0,gh)$, so G is a subgroup.

Let c be a 1-cochain with $dc(e,g) = dc(g,e) = 0$, the latter property occurring if and only if $c(e) = 0$. Consider the map $\psi_{f,c} \colon \mathscr{E}_f \to \mathscr{E}_{f+dc}$ given by

$$\psi_{f,c}(a,g) = (a - c(g), g).$$

We have

$$\begin{aligned}\psi_{f,c}(a+gb+f(g,h),gh) &= (a+gb+f(g,h)-c(gh),gh)\\ &= (a+gb+f(g,h)+dc(g,h)-gc(h)+c(g),gh) = \psi_c(a,g)\psi_c(a,h),\end{aligned}$$

so ψ_c is a homomorphism, and it is clearly has inverse $\psi_{f+dc,-c}$. Thus, we have a well-defined map from $H^2(G,A)$ to $\mathscr{E}(G,A)$.

It remains to construct an inverse, which we sketch as the computations all follow from what we have already done. Given a group extension, we indeed always have a 2-cochain $f(g,h) \colon G^2 \to A$ defining the multiplication. We claim that f is a factor set. For this, associativity again tells us that f is a 2-cocycle, and the fact that $(0,e)$ is a two-sided identity forces $f(e,g) = f(g,e) = 0$ for all $g \in G$. The resulting association is clearly inverse on the level of extensions and cochains. If $\theta \colon \mathscr{E} \to \mathscr{E}'$ is an isomorphism of group extensions of G by A, then $\theta(0,g) = (-c(g),g)$ for some $c \colon G \to A$ that has the property that if the factor set is associated to $\mathscr{E}$ is dc plus the factor set associated to $\mathscr{E}'$. □

REMARK 12.9.23. Theorem 12.9.22 tells us that $\mathscr{E}(G,A)$ also has a group structure, which may also be given an explicit description. Given E and E' extensions of G by A, their product is

$$E * E' = (E \times_G E')/\langle (a,-a) \mid a \in A\rangle.$$

This product is known as the *Baer sum* of the two extensions.

Let's give a group-theoretic application of this description of $H^2(G,A)$.

PROPOSITION 12.9.24 (Schur). *Let G be a group of order mn, where m and n are relatively prime positive integers. Then every abelian normal subgroup of order n has a complement in G of order m.*

PROOF. Let N be an abelian normal subgroup of G, and set $H = G/N$. Let $f \colon G^2 \to N$ be a factor set corresponding to G as an extension of H by N by Theorem 12.9.22. For every $h \in H$, let

$$t(h) = \prod_{k \in H} f(h,k),$$

which makes sense as H is abelian. For $h, h' \in H$, we have

$$\prod_{k \in H} f(h, h'k) = \prod_{k \in H} f(h,k) = t(h).$$

Now

$$f(h,h')^m t(hh') = \prod_{k \in H} f(h,h')^m f(hh',k) = \prod_{k \in H} hf(h',k)h^{-1} f(h,h'k) = ht(h')h^{-1} \cdot t(h).$$

Let $a, b \in \mathbb{Z}$ be such that $am + bn = 1$. We then have

$$f(h,h') = f(h,h')^{am+bn} = dt(h,h')^a.$$

Thus, f is a coboundary, so G is a split extension by Theorem 12.9.22. In particular, it contains a subgroup of order n, isomorphic to H. □

REMARK 12.9.25. Though we do not prove it, we have $H^i(G,A) = 0$ for all $i \geq 1$ whenever G and A are finite of relatively prime order. In fact, for any finite group G and G-module A, the exponent of $H^i(G,A)$ divides the order of G.

DEFINITION 12.9.26. A *Hall subgroup* of a finite group is a subgroup with relatively prime order and index.

We can extend Proposition 12.9.24 from abelian to arbitrary normal subgroups.

THEOREM 12.9.27 (Schur-Zassenhaus). *Every normal Hall subgroup of a finite group has a complement.*

PROOF. Let N be a normal Hall subgroup of G of order n and index m. If N is abelian, then the result follows from Proposition 12.9.24. Suppose the result holds true in the case of normal subgroups of order less than $n \geq 2$. Let p be a prime dividing n. Let P be a Sylow p-subgroup of G. Then PN/N has p-power order dividing m. Since m and n are relatively prime, this forces P to be contained in N. In other words, the Sylow p-subgroups of N and G are the same. Now

$$[G : N_G(P)] = n_p(G) = n_p(N) = [N : N_N(P)],$$

so $[N_G(P) : N_N(P)] = m$. On the other hand, $N_N(P)/P$ has order prime to m and less than n, being properly contained in N/P. Furthermore, $N_N(P) = N \cap N_G(P)$ is normal in $N_G(P)$. By induction on n, we have that there exists a subgroup K of $N_G(P)$ with K/P isomorphic to $N_G(P)/N_N(P)$ and $|K/P| = m$.

Since P is a p-group, its center $Z = Z(P)$ is nontrivial. It is also a characteristic subgroup of P, so it is normal in K. By induction, P/Z has a has a complement in K/Z, equal to H/Z for some subgroup H of K, which necessarily has order m. This group H is the desired complement to N. □

12.10. Galois cohomology

We briefly consider the cohomology of finite Galois extensions. We have the following generalization of Hilbert's Theorem 90, which also has the same name.

THEOREM 12.10.1 (Hilbert's Theorem 90). *Let L/K be a finite Galois extension with Galois group G. Then $H^1(G, K^\times) = 0$.*

PROOF. Let $f \colon G \to L^\times$ be a 1-cocycle. We view the elements $\sigma \in G$ as abelian characters $L^\times \to L^\times$. As distinct characters of $L^\times$, these characters form a linearly independent set. The sum $\sum_{\sigma\in G} f(\sigma)\sigma$ is therefore a nonzero map $L^\times \to L$. Let $\alpha \in L^\times$ be such that $z = \sum_{\sigma\in G} f(\sigma)\sigma(\alpha) \neq 0$. For any $\tau \in G$, we have

$$\begin{aligned}\tau^{-1}(z) &= \sum_{\sigma\in G} \tau^{-1}(f(\sigma)) \cdot \tau^{-1}\sigma(\alpha) = \sum_{\sigma\in G} \tau^{-1}(f(\tau\sigma))\sigma(\alpha) \\ &= \sum_{\sigma\in G} \tau^{-1}(f(\tau) \cdot \tau f(\sigma))\sigma(\alpha) = \tau^{-1}(f(\tau)) \sum_{\sigma\in G} f(\sigma)\sigma(\alpha) = \tau^{-1}(f(\tau))z.\end{aligned}$$

Thus,

$$f(\tau) = \frac{z}{\tau(z)},$$

so f is the 1-coboundary of z^{-1}. □

To see how this implies Hilbert's theorem 90 in the case of finite cyclic extensions, we prove the following result on the cohomology of cyclic groups.

PROPOSITION 12.10.2. *Let G be a finite cyclic group and A be a $\mathbb{Z}[G]$-module. Then for $i \geq 1$, we have*

$$H^i(G,A) \cong \begin{cases} A^G/N_GA & \text{if } i \text{ is even,} \\ A[N_G]/I_GA & \text{if } i \text{ is odd,} \end{cases}$$

where $A[N_G]$ is the kernel of multiplication by N_G on A.

PROOF. Let g be a generator of G, and consider the augmented resolution of $\mathbb{Z}$ given by

$$\cdots \to \mathbb{Z}[G] \xrightarrow{N_G} \mathbb{Z}[G] \xrightarrow{g-1} \mathbb{Z}[G] \xrightarrow{N_G} \mathbb{Z}[G] \xrightarrow{g-1} \mathbb{Z}[G] \xrightarrow{\varepsilon} \mathbb{Z} \to 0,$$

where ε is the augmentation maps. Note that $\mathrm{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G],A) \cong A$ by evaluation at 1, and the map N_G (resp., $g-1$) on $\mathbb{Z}[G]$ induces N_G (resp., $g-1$) on A via these isomorphisms. The groups $H^i(G,A)$ are then the cohomology groups of the complex

$$A \xrightarrow{g-1} A \xrightarrow{N_G} A \xrightarrow{g-1} A \to \cdots,$$

which have the desired form. □

REMARK 12.10.3. Suppose that L/K is finite cyclic with Galois group G having generator σ. Proposition 12.10.2 implies that

$$H^1(G,K^\times) \cong \frac{\ker N_{L/K}}{\{\sigma(\alpha)/\alpha \mid \alpha \in L^\times\}},$$

which is trivial by Theorem 12.10.1. This is exactly the statement of Hilbert's Theorem 90 for finite cyclic extensions.

We next see how we can use Galois cohomology to study Kummer theory.

PROPOSITION 12.10.4. *Suppose that L/K is a finite extension with Galois group G. Let n be a positive integer not divisible by the characteristic of K. Then there is an isomorphism*

$$H^1(G,\mu_n) \cong \frac{K^\times \cap L^{\times n}}{K^{\times n}},$$

where the class of an element $a \in K^\times \cap L^{\times n}$ in the quotient corresponds to the class of the cocycle

$$\chi_a(\sigma) = \frac{\sigma(\alpha)}{\alpha},$$

where $\alpha \in L^\times$ with $\alpha^n = a$

PROOF. The short exact sequence

$$1 \to \mu_n(L) \to L^\times \xrightarrow{n} L^{\times n} \to 1$$

of G-modules gives rise to a long exact sequence

$$0 \to \mu_n(K) \to K^\times \xrightarrow{n} K^\times \cap L^{\times n} \xrightarrow{\partial} H^1(G, \mu_n) \to 0,$$

where Hilbert's Theorem 90 gives the final equality. That $a \in K^\times \cap L^{\times n}$ is sent to the class of χ_a follows from the definition of the connecting homomorphism by the Snake lemma. □

This leads to the following definition.

DEFINITION 12.10.5. Let n be a positive integer not divisible by the characteristic of K. For $a \in K^\times$, a *Kummer cocycle* attached to K is a map $\chi_a \colon G_K \to \mu_n$ given by

$$\chi_a(\sigma) = \frac{\sigma(\alpha)}{\alpha},$$

where $\alpha \in (K^{\mathrm{sep}})^\times$ with $\alpha^n = a$.

REMARK 12.10.6. The Kummer cocycle χ_a in Definition 12.10.5 actually depends on the choice of nth root of a up to a 1-coboundary of an element of μ_n. If $\mu_n \subseteq K$, however, it is unique and is the Kummer character of a. In this case, Proposition 12.10.4 reduces to

$$\mathrm{Hom}(G, \mu_n) \cong \Delta / K^{\times n},$$

where $\Delta = K^\times \cap L^{\times n}$. This in turn yields the perfect pairing of Kummer duality.

We next turn to the question of the structure of $H^2(G, L^\times)$ for a finite Galois extension L/K with Galois group G. We fix such an extension L/K with Galois group G in what follows.

DEFINITION 12.10.7. A *central simple algebra* over a field K is a simple K-algebra with center equal to K.

EXAMPLE 12.10.8. Any matrix algebra $M_n(D)$ over a division algebra D is a central simple algebra over the center $Z(D)$, which is a field. For instance, if $\mathbb{H}$ denotes the ring of quaternions, then $M_n(\mathbb{H})$ is a central simple $\mathbb{R}$-algebra.

PROPOSITION 12.10.9. *Let $f \in Z^2(G, L^\times)$ be a factor set. Let B_f be an L-vector space with basis b_σ for $g \in G$. Define a multiplication on B_f as the unique binary operation extending the scalar multiplication $L \times B_f \to B_f$ and satisfying*

$$b_\sigma \alpha = \sigma(\alpha) b_\sigma \quad \text{and} \quad b_\sigma b_\tau = f(\sigma, \tau) b_{\sigma\tau}.$$

for $\sigma, \tau \in G$ and $\alpha \in L$. Then B_f is a central simple K-algebra with identity b_1.

PROOF. We have

$$\begin{aligned}(b_\sigma b_\tau) b_\rho = f(\sigma,\tau)(b_{\sigma\tau} b_\rho) = f(\sigma,\tau) f(\sigma\tau,\rho) b_{\sigma\tau\rho} &= \sigma(f(\tau,\rho)) f(\sigma,\tau\rho) b_{\sigma\tau\rho} \\ &= \sigma(f(\tau,\rho))(b_\sigma b_{\tau\rho}) = b_\sigma(f(\tau\rho) b_{\tau\rho}) = b_\sigma(b_\tau b_\rho).\end{aligned}$$

It follows that that B_f is an associative L-algebra with K in its center. Note that $b_1 = 1$ in the ring B_f since $f(\sigma, 1) = f(1, \sigma) = 1$ for all $\sigma \in G$.

Let $\beta \in L^\times$ generate L/K. Let

$$z = \sum_{\sigma \in G} \alpha_\sigma b_\sigma \in Z(G).$$

. Then $z\beta = \beta z$, so

$$\sum_{\sigma \in G} (\alpha_\sigma \sigma(\beta) - \beta \alpha_\sigma) b_g = 0,$$

and therefore $\sigma(\beta) = \beta$ for all $\sigma \in G$ with $\alpha_\sigma = 0$. Since β is a generator of L/K, this forces $\alpha_\sigma = 0$ for all $\sigma \neq 1$, so $z = \alpha_1 \in K$. Thus, $Z(B_f) = K$.

Next, let I be a nonzero ideal of B_f, and let $x \in B_f$ be an element with a minimal number k of nonzero coefficients in its expression as an L-linear combination of elements of G. If σ and τ are distinct elements of G for which x has nonzero coefficients, then $\sigma(\beta) \neq \tau(\beta)$. Then $x - \tau(\beta)x\beta^{-1} \in I$, but its b_τ-coefficient is now zero, while its b_σ-coefficient is not, and it has no nonzero coefficients that x does not have. This contradicts the minimality of k, forcing it to be 1. Thus, $x = \alpha b_\sigma$ for some $\alpha \in L^\times$ and $\sigma \in G$. But such an x is a unit in B_f, so $I = B_f$. Thus, B_f is a simple ring. □

DEFINITION 12.10.10. For a factor set $f \colon G \to L^\times$, the K-algebra B_f of Proposition 12.10.9 is the *crossed product algebra* of f.

CHAPTER 13

Representation theory

13.1. Semisimple modules

The following definitions will be of special interest in the case of a group ring over a field.

DEFINITION 13.1.1. A module M over a ring R is *simple*, or *irreducible*, if it has no nonzero, proper R-submodules. Otherwise, M is said to be *reducible*.

DEFINITION 13.1.2. A module M over a ring R is *indecomposable* if it is not the direct sum of two proper submodules.

DEFINITION 13.1.3. A module M over a ring R is *semisimple*, or *completely reducible*, if it is a direct sum of irreducible submodules.

REMARK 13.1.4. By definition, a module is simple if and only if it is both semisimple and indecomposable.

EXAMPLES 13.1.5.

a. Any division ring D is simple as a left module over itself, as it has no nontrivial left ideals.

b. Any vector space V over a field F is semisimple as an F-module, in that it has a basis that allows us to express it (up to isomorphism) as a direct sum of copies of F.

c. The ring $\mathbb{Z}$ is indecomposable as a $\mathbb{Z}$-module, but it is not simple, as it contains proper, nontrivial submodules $n\mathbb{Z}$ for $n \geq 2$.

d. Any simple $\mathbb{Z}$-module is a simple abelian group, so isomorphic to $\mathbb{Z}/p\mathbb{Z}$ for some prime p.

e. The $\mathbb{Z}$-module $\mathbb{Z} \oplus \mathbb{Z}$ is neither semisimple nor indecomposable, as it is not a direct sum of simple $\mathbb{Z}$-modules.

f. Let R be the ring of upper-triangular matrices in $M_2(F)$ for F a field, and consider the R-module $M = F^2$ under left multiplication of column vectors. Then M has a simple submodule $N = F \cdot e_1$, so M is not simple. Moreover, M is not semisimple, as $M = F \cdot v$ for any $v \notin N$, so N has no complement in M.

Semisimple modules have the following equivalent characterizations.

PROPOSITION 13.1.6. *Let M be an R-module. The following are equivalent:*

i. M is semisimple.

ii. M is a sum of simple submodules.

iii. Every submodule of M is a direct summand.

PROOF. That (i) implies (ii) is clear. As for (ii) implies (iii), let N be a submodule of

$$M = \bigoplus_{i \in I} M_i,$$

where the M_i are simple. Then $N \cap M_i$ is either 0 or M_i for each i, and N is the direct sum of the M_i for which $N \cap M_i = M_i$.

That (iii) implies (i) is proven as follows. We first claim that any nonzero R-module M contains a nonzero simple submodule. To see this, choose $m \in M$, and replace M with Rm without loss of generality. Let N be a maximal R-submodule of M not containing n, which exists by Zorn's Lemma. Then $M = N \oplus N'$ for some nonzero R-submodule N'. Now N' must be simple, since any $Q \subseteq N'$ has $N \oplus Q$ containing a and therefore equals M.

Now consider the nonempty set X of semisimple submodules of M under inclusion. The union of any chain C in X is semisimple (as the reader may check), so X has a maximal element N by Zorn's lemma. Let N' be a complement to N in M, so $M = N \oplus N'$. If N' is nonzero, then N' contains a simple submodule Q by the claim, and $N \oplus Q$ is semisimple, contradicting the maximality of M. So, $M = N$ is semisimple. □

We define semisimple rings in a manner that does not obviously relate to simple rings.

DEFINITION 13.1.7. A nonzero ring is *semisimple* if it is semisimple as a left module over itself.

The following contains equivalent conditions for a ring to be semisimple.

THEOREM 13.1.8. *The following conditions on a nonzero ring R with unity are equivalent:*

i. R is semisimple,

ii. every R-module is semisimple,

iii. every R-module is projective,

iv. every R-module is injective.

PROOF. Suppose that R is semisimple, and let M be an R-module. Then M is a sum of its cyclic submodules, so by Lemma 13.1.6, it suffices to see that quotients Q of R are semisimple. Again employing Lemma 13.1.6, the kernel I of the quotient map $R \to Q$ is a direct summand, so Q is isomorphic to a left ideal of R, which is semisimple as R is.

That (ii) implies (iii) is an immediate consequence of Lemma 13.1.6. Every surjection of R-modules is split if and only if every injection of R-modules is split by Proposition 12.4.3, so (iii) and (iv) are equivalent, noting Proposition 12.4.8 and Lemma 12.4.13. Finally, (iv) tells us that every R-submodule of an R-module is a direct summand, so is semisimple by Lemma 13.1.6. □

We claim that simple rings are indeed semisimple, so long as we assume that descending chains of left or right ideals terminate. This can be seen directly for matrix rings over division rings, using Morita equivalence.

DEFINITION 13.1.9. A ring R is *left artinian* (resp., *right artinian* if it satisfies the descending chain condition on left ideals (resp., right ideals).

LEMMA 13.1.10. *If a ring R with unity is the sum of a collection of its nonzero left ideals, then it is also a sum of a finite subcollection.*

PROOF. If $\{I_x \mid x \in X\}$ is a set of nonzero left ideals of R such that $R \cong \sum_{x\in X} I_x$ as left-modules, then we can write $1 = \sum_{j=1}^{n} a_j$ for some $n \geq 1$, where $a_j \in I_{x_j}$ for some $x_j \in I$. But then the left ideals I_{x_j} with $1 \leq j \leq n$ generate R as a left R-module. □

COROLLARY 13.1.11. *A semisimple ring R is left artinian, isomorphic as an R-module to the direct sum of its finitely many minimal left ideals.*

PROOF. By definition, R is isomorphic to the direct sum of its minimal ideals. Since the sum is direct, no proper subcollection of the minimal ideals generates R. Lemma 13.1.10 then tells us that the collection of minimal ideals must be finite. It follows that R is left artinian. □

PROPOSITION 13.1.12. *Let R be a left (or right) artinian simple ring. Then R is semisimple.*

PROOF. Let R be left artinian and simple. We first claim that R has a simple R-submodule (i.e., left ideal). For this, construct a possibly finite sequence of left ideals J_i of R recursively, starting with $J_1 = R$, and then for $i \geq 1$, taking J_{i+1} to be a proper simple submodule if J_i is not simple. Since R is left artinian, we must have that that the process terminates, so R has a simple submodule.

Now, consider the nonzero sum M of all distinct simple R-submodules of R. Let N be a simple submodule of R, and let $r \in R$. Then Nr is isomorphic to a quotient of n, so is either 0 or simple. In particular, Nr is contained in M, and therefore $Mr \subseteq M$. Thus, M is not only a left ideal of R, but a right ideal as well, and therefore $M = R$.

By Lemma 13.1.10, the R-module R is then a finite sum of distinct simple left ideals: say R is the sum of N_i simple for $1 \leq i \leq k$, where k is minimal. If the intersection N_i with the sum M_i of the N_j for $v \neq i$ is nonzero, then it must equal N_i, as N_i is simple. But then $N_i \subseteq M_i$, so $M_i = R$, which contradicts the minimality of k. So, R is in fact the direct sum of the N_i, as required. □

The following is an easy but very useful fact regarding homomorphisms of simple modules.

LEMMA 13.1.13 (Schur's lemma). *Let R be a ring, and let M and N be simple R-modules. Then any nonzero homomorphism $f\colon M \to N$ is an isomorphism.*

PROOF. The kernel of f is a proper R-submodule of M, hence zero, and the image of f is a nonzero R-submodule of N, hence N. Thus f is bijective. □

Since every nonzero R-linear endomorphism of a simple module is invertible by Schur's lemma, we have the following corollary.

LEMMA 13.1.14. *Let R be a nonzero ring. The ring* $\mathrm{End}_R(M)$ *of R-linear endomorphisms of a simple module M is a division ring.*

Recall that a ring is simple if it has no nonzero ideals, and by Remark 3.9.8, matrix rings over division algebras are simple. We have the following consequence of Schur's lemma

LEMMA 13.1.15. *Let M be a simple R-module, and let $n \geq 1$. Then* $\mathrm{End}_R(M^n) \cong M_n(D)$, *where D is the division ring* $\mathrm{End}_R(M)$. *In particular,* $\mathrm{End}_R(M^n)$ *is a simple ring.*

PROOF. We define a homomorphism

$$\Phi\colon M_n(D) \xrightarrow{\sim} \mathrm{End}_R(M^n)$$

on a matrix $C = (\phi_{ij}) \in M_n(D)$ by

$$\Phi(C)(m_1,\dots,m_n) = \left(\sum_{j=1}^{n} \phi_{1j}(m_j),\dots,\sum_{j=1}^{n} \phi_{nj}(m_j)\right).$$

Every endomorphism $\phi \in \mathrm{End}_R(M^n)$ is determined uniquely by the collection of maps $\phi_{ij} = \pi_j \circ \phi \circ \iota_i \in \mathrm{End}_R(M)$, where π_i and ι_i denote the ith projection and inclusion maps, so this is one-to-one and onto. □

We can improve this lemma to treat a finite direct sum of arbitrary simple modules.

LEMMA 13.1.16. *Let R be a nonzero ring. Let M be an R-module that is isomorphic to a direct sum $N_1^{n_1} \oplus N_2^{n_2} \oplus \cdots \oplus N_k^{n_k}$ with the N_i mutually nonisomorphic simple modules and $n_i \geq 1$ for $1 \leq i \leq k$. Then we have an isomorphism of rings*

$$\mathrm{End}_R(M) \cong \prod_{i=1}^{k} M_{n_i}(D_i),$$

where D_i is the division ring $\mathrm{End}_R(N_i)$.

PROOF. Let $\pi_i\colon M \to N_i^{n_i}$ and $\iota_i\colon N_i^{n_i} \to M$ be the projection and inclusion maps. Any R-module endomorphism f of M determines and is determined by the homomorphisms $f_{i,j} = \pi_j \circ f \circ \iota_i\colon N_i^{n_i} \to N_j^{n_j}$ for $1 \leq i, j \leq n$. But $\mathrm{Hom}_R(N_i,N_j) = 0$ for $i \neq j$, so $\mathrm{Hom}_R(N_i^{n_i},N_j^{n_j}) = 0$ for $i \neq j$ as well. Therefore, the product of restriction maps to $N_i^{n_i}$ yields the first of the isomorphisms in

$$\mathrm{End}_R(M) \cong \prod_{i=1}^{k} \mathrm{End}_R(N_i^{n_i}) \cong \prod_{i=1}^{k} M_{n_i}(D_i),$$

where the second isomorphism is by Lemma 13.1.15 □

Evaluation at 1 gives the isomorphism in the following lemma.

LEMMA 13.1.17. *We have* $\mathrm{End}_R(R) \to R^{\mathrm{op}}$ *as rings.*

We now come to the Artin-Wedderburn theorem, which classifies semisimple rings.

THEOREM 13.1.18 (Artin-Wedderburn theorem). *A nonzero ring is semisimple if and only if it is isomorphic to a direct product of matrix algebras over division rings.*

PROOF. For any nonzero ring R, we have an isomorphism

$$R^{\mathrm{op}} \xrightarrow{\sim} \mathrm{End}_R(R), \qquad r \mapsto (s \mapsto sr).$$

Supposing that R is semisimple, we have by Corollary 13.1.11 that $R \cong N_1^{n_1} \oplus N_2^{n_2} \oplus \cdots \oplus N_k^{n_k}$ with the N_i mutually nonisomorphic simple left R-modules and $n_i \geq 1$ for $1 \leq i \leq k$. Noting Lemma 13.1.17, we then have

$$R^{\mathrm{op}} \cong \mathrm{End}_R(R) \cong \prod_{i=1}^{k} M_{n_i}(D_i),$$

where $D_i = \mathrm{End}_R(N_i)$ is a division ring. By taking the opposite ring of both sides, we obtain

$$R \cong \prod_{i=1}^{k} M_{n_i}(D_i)^{\mathrm{op}} = \prod_{i=1}^{k} M_{n_i}(D_i^{\mathrm{op}}),$$

and D_i^{op} is a division ring as well.

On the other hand, suppose that $R \cong \prod_{i=1}^{k} M_{n_i}(E_i)$ for some division algebras E_i. The matrix rings $M_{n_i}(E_i)$ are semisimple left modules over $M_{n_i}(E_i)$, isomorphic to a direct sum of the simple submodules of column vectors. They are also then semisimple as modules for the larger ring R, since the action of R on $M_{n_i}(E_i)$ by left multiplication factors through $M_{n_i}(E_i)$. Thus, R is a semisimple ring as a direct sum of these as a left R-module. □

Here are some corollaries. The first follows directly from Proposition 13.1.12 and the Artin-Wedderburn theorem.

COROLLARY 13.1.19. *A nonzero ring is left artinian and simple if and only if it is isomorphic to a matrix ring over a division ring.*

Consequently, we have the following, which explains the relationship between simple and semisimple rings.

COROLLARY 13.1.20. *A nonzero ring is semisimple if and only if it is isomorphic to a finite direct product of left artinian simple rings.*

For algebras over a field, we obtain Wedderburn's theorem.

COROLLARY 13.1.21 (Wedderburn). *An algebra over a field F is semisimple if and only if it is a product of finite-dimensional simple F-algebras, and these simple algebras are isomorphic to matrix rings over finite-dimensional division algebras over F.*

The following greatly limits the choice of finite-dimensional division algebras over algebraically closed fields.

PROPOSITION 13.1.22. *Let D be a finite-dimensional division algebra over an algebraically closed field F. Then $D = F$.*

PROOF. Let $\gamma \in D$. Note that γ commutes with every element of F, so $F(\gamma)$ is a field. Since D is finite-dimensional over F, the elements γ^i for $i \geq 0$ are linearly dependent over F, and therefore γ is algebraic over F. Thus $F(\gamma) = F$, which is to say $\gamma \in F$. □

COROLLARY 13.1.23. *Let A be a finite-dimensional, semsimple F-algebra, where F is an algebraically closed field. Then A is isomorphic to a direct product of matrix algebras with F-entries.*

For commutative rings, we have this:

COROLLARY 13.1.24. *A commutative semisimple ring is a finite direct product of fields. A finite-dimensional commutative semisimple algebra over a field F is isomorphic to a direct product of finite field extensions of F.*

DEFINITION 13.1.25. Let R be a ring. An *idempotent* in R is a nonzero element $e \in R$ such that $e^2 = e$.

DEFINITION 13.1.26. Let R be a ring. We say two idempotents $e, f \in R$ are *orthogonal* if $ef = fe = 0$.

REMARK 13.1.27. Any finite sum of orthogonal idempotents is also an idempotent.

DEFINITION 13.1.28. We say that an idempotent e in a ring R is *primitive* if eR is a subring of R that is not a product of two subrings of R.

LEMMA 13.1.29. *Let R be a nonzero ring and $k \geq 1$. Then $R = R_1 \times R_2 \times \cdots \times R_k$ with R_i rings for $1 \leq i \leq k$ if and only if there exist mutually orthogonal idempotents $e_1, e_2, \ldots, e_k$ in $Z(R)$ such that $e_1 + e_2 + \cdots + e_k = 1$. These may be chosen so that $R_i = (e_i)$ in R.*

PROOF. If $R = \prod_{i=1}^k R_i$, then let e_i be the identity in R_i. The e_i are then clearly mutually orthogonal, central idempotents. Set $e = \sum_{i=1}^k e_i$. If $1 = (r_1, r_2, \ldots, r_k) \in R$, then

$$e = e \cdot 1 = (e_1 r_2, e_2 r_2, \ldots, e_k r_k) = r.$$

Conversely, given $e_1, e_2, \ldots, e_k$, set $R_i = Re_i$ in R. For any $r \in R$, we have $r = \sum_{i=1}^k re_i$, so $R = \sum_{i=1}^k R_i$. If $r_i \in R_i$ for each i, then set $r = \sum_{i=1}^k r_i$. This satisfies $re_j = r_j$ for each $1 \leq j \leq k$, so $r = 0$ if and only if each $r_j = 0$, and thus $R = \bigoplus_{i=1}^k R_i$ as left R-modules. Since each e_i is central in R, each R_i is also a right ideal and a ring with unit element e_i, so this decomposition is actually as a product of subrings with unity. □

EXAMPLE 13.1.30. If R is a direct product of matrix rings, then it has a set of mutually orthogonal idempotents consisting of the identity matrices in those rings.

LEMMA 13.1.31. *Let $R = \prod_{i=1}^k R_i$ be a direct product of rings R_i, and let e_i be identity element of R_i. Let M be a left R-module. Then $M = \bigoplus_{i=1}^k e_i M$ as an R-module.*

PROOF. Any $m \in M$ can be written as $m = e_1 m + e_2 m + \cdots + e_k m$, so $M = \sum_{i=1}^k e_i M$. If $m_1 + m_2 + \cdots + m_k = m$ with $m_i \in e_i M$ for $1 \leq i \leq k$, then

$$m_i = e_i(m_1 + m_2 + \cdots + m_k) = e_i m$$

for each i, so the representation of m as an element of the sum is unique. Therefore, M is the direct sum of the $e_i M$. □

13.2. Representations of groups

Let G be a group.

PROPOSITION 13.2.1. *Let R be a commutative ring, let G be a group, and let M be an R-module. There is a one-to-one correspondence between* *i. homomorphisms $\rho\colon G \to \mathrm{Aut}_R(M)$,*

ii. $R[G]$-module structures given by R-bilinear maps $\phi\colon R[G] \times M \to M$

such that ρ corresponds to a unique ϕ with $\rho(g)(m) = \phi(g,m)$ for all $g \in G$ and $m \in M$.

PROOF. If $\rho\colon G \to \mathrm{Aut}_R(M)$ is a homomorphism, then we define

$$\left(\sum_{g\in G} a_g g\right) \cdot m = \sum_{g\in G} a_g \rho(g)(m) \tag{13.2.1}$$

for $\sum_{g\in G} a_g g \in R[G]$ and $m \in M$. For a fixed m, this provides the unique $R[G]$-module homomorphism $R[G] \to M$ that sends g to $\rho(g)(m)$ by the R-freeness of $R[G]$. In other words, the operation $R[G] \times M \to M$ is left distributive. Since $\rho(g)(m+m') = \rho(g)(m) + \rho(g)(m')$ for $g \in G$ and $m, m' \in M$ in that $\rho(g) \in \mathrm{Aut}_R(M)$, right distributivity follows from the definition in (13.2.1) as well.

Conversely, given an $R[G]$-module M, we define $\rho\colon G \to \mathrm{Aut}_R(M)$ by $\rho(g)(m) = g \cdot m$ for $g \in G$ and $m \in M$. Note that

$$\rho(g)(m+m') = g(m+m') = gm + gm' = \rho(g)(m) + \rho(g)(m')$$

and $\rho(g)(rm) = g(rm) = r(gm) = r\rho(g)(m)$, so $\rho(g)$ is indeed an element of $\mathrm{Aut}_R(M)$. Moreover,

$$\rho(gg')(m) = (gg')m = g(g'm) = \rho(g)(g'm) = \rho(g)(\rho(g')(m)) = (\rho(g) \circ \rho(g'))(m),$$

so ρ is a homomorphism. □

REMARK 13.2.2. To give an $R[G]$-module structure on an R-module M, it suffices to give an operation $G \times M \to M$ such that the map $g\colon M \to M$ defined by left multiplication is R-linear.

REMARK 13.2.3. The trivial $R[G]$-module R on which $g \cdot r = r$ for all $g \in G$ and $r \in R$ corresponds to the homomorphism $\rho\colon G \to \mathrm{Aut}_R(R) \cong R^\times$ with $\rho(g) = 1$ for all $g \in G$.

EXAMPLE 13.2.4. Let R be a commutative ring and G be a group. We may view $R[G]$ as a left $R[G]$-module under left multiplication. This corresponds to the homomorphism $\rho\colon G \to \mathrm{Aut}_R(R[G])$ that takes g to left multiplication by g on $R[G]$.

We now focus on the special case that R is a field, which yields group representations. From now on in this section, we let F denote a field.

DEFINITION 13.2.5. A *representation*, or *group representation*, of a group G over a field F is an F-vector space V, together with a homomorphism $\rho\colon G \to \mathrm{Aut}_F(V)$. We also say that V is an *F-representation* of G.

REMARK 13.2.6. By Proposition 13.2.1, to make an F-vector space V into an $F[G]$-module V is equivalent to providing a homomorphism $\rho\colon G \to \mathrm{Aut}_F(V)$ that makes it into a representation of G.

DEFINITION 13.2.7. We say that a representation $\rho\colon G \to \mathrm{Aut}_F(V)$ is *finite-dimensional* if V is a finite-dimensional F-vector space, in which case $\dim_F V$ is its *dimension*, also known as its *degree*.

Representations form one of the most important tools in the study of the structure of groups.

EXAMPLE 13.2.8. Let G be a subgroup of $\mathrm{GL}_n(F)$. Then the inclusion $\rho\colon G \to \mathrm{GL}_n(F)$ defines a representation of G, and this turns F^n into an $F[G]$-module, where $g \in G$ acts on $v \in F^n$ by left multiplication of the column vector v by the matrix corresponding to g.

EXAMPLE 13.2.9. The representation $\rho\colon \mathbb{R} \to \mathrm{GL}_2(\mathbb{R})$ given by

$$\rho(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

is a two-dimensional real representation of the additive group $\mathbb{R}$.

DEFINITION 13.2.10.

a. The *trivial representation* of G over F is F with the trivial G-action.

b. The *regular representation* of G over F is $F[G]$ with the action of $F[G]$ on itself by left multiplication.

REMARK 13.2.11. Two F-representations V and W of G are isomorphic if V and W are isomorphic as $F[G]$-modules. Phrased in terms of the corresponding homomorphisms ρ_V and ρ_W, this says that ρ_V and ρ_W are conjugate by the isomorphism $\varphi\colon V \to W$: that is, $\rho_W(g) = \varphi \circ \rho_V(g) \circ \varphi^{-1}$ for all $g \in G$.

EXAMPLES 13.2.12. Let V and W be F-representations of a group G.

a. The F-vector space $V \otimes_F W$ is a representation of G with respect to the diagonal G-action $g \cdot (v \otimes w) = gv \otimes gw$ for $g \in G$, $v \in V$ and $w \in W$.

b. The F-vector space $\mathrm{Hom}_F(V,W)$ is a representation of G with respect to the G-action $(g \cdot \varphi)(v) = g\varphi(g^{-1}v)$ for $g \in G$, $\varphi \in \mathrm{Hom}_F(V,W)$, and $v \in V$.

As a special case, we have the following.

DEFINITION 13.2.13. Let V be an F-representation of a group G. The *dual representation* to V is $V^* = \mathrm{Hom}_F(V,F)$.

The reader will easily check the following.

LEMMA 13.2.14. *Let V and W be F-representations of a group G, and suppose that W is finite-dimensional. Then* $\mathrm{Hom}_F(V,W) \cong V^* \otimes_F W$.

TERMINOLOGY 13.2.15. We speak of F-representations of a group G as being simple, indecomposable, and so forth, if the $F[G]$-modules that define them have these properties.

DEFINITION 13.2.16. An F-representation V of a group G is called *faithful* if $\rho_V\colon G \to \mathrm{Aut}_F(V)$ is injective.

DEFINITION 13.2.17. A *subrepresentation* W of an F-representation V of a group G is an $F[G]$-submodule of V.

REMARK 13.2.18. An irreducible (i.e., simple) representation is one that has no nonzero, proper subrepresentations.

EXAMPLES 13.2.19.

a. All one-dimensional representations of a group are irreducible.

b. Let $D_p = \langle r, s\rangle$ be the dihedral group of prime order p, and let $\rho\colon D_p \to \mathrm{GL}_2(\mathbb{F}_p)$ be the representation with $\rho(s) = \left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $\rho(r) = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Then ρ is indecomposable but not irreducible, since the $\mathbb{F}_p$-submodule W of $V = \mathbb{F}_p^2$ spanned by e_1 is left stable by (i.e., is closed under) the action of D_p, so is a subrepresentation. On the other hand, W does not have a complement in V (i.e., the only line in $\mathbb{F}_p^2$ that is stabilized by D_p is W).

c. The regular representation of a finite group is faithful, whereas the trivial representation is not faithful unless the group is trivial.

Let us rephrase Schur's lemma in the context of representations.

LEMMA 13.2.20. *Let V be an irreducible F-representation of G. Then $\mathrm{End}_{F[G]}(V)$ is a division algebra over F.*

PROOF. This is an immediate consequence of Lemma 13.1.14, noting that the the endomorphisms given by multiplication by elements of F are contained in the center of $\mathrm{End}_{F[G]}(V)$. □

By Proposition 13.1.22, this has the following corollary.

COROLLARY 13.2.21. *Let V be a finite-dimensional irreducible F-representation of a group G, where F is algebraically closed. Then $\mathrm{End}_{F[G]}(V) \cong F$.*

DEFINITION 13.2.22. Let V and W be representations of G over a field F, with V semisimple and W irreducible. The *multiplicity* of W in V is the largest nonnegative integer n such that W^n is isomorphic to a subrepresentation of V. We say that W *occurs with multiplicity n* in V.

LEMMA 13.2.23. *Let V be an F-representation of a finite group G. Let E/F be a field extension. Then $E \otimes_F V$ is an $E[G]$-module under the action $g \cdot (\alpha \otimes v) = \alpha \otimes gv$ with the same character as V.*

PROOF. Note that $E[G] \cong E \otimes_F F[G]$, and the action described is just the usual action of a tensor product of algebras on a tensor product of modules over them. □

DEFINITION 13.2.24. For an F-representation V of a group G and a field extension E/F, the E-representation $E \otimes_F V$ is called the *base change* of V from F to E.

13.3. Maschke's theorem

In this section, we let G be a finite group, and we let F be a field of characteristic not dividing the order of G.

THEOREM 13.3.1 (Maschke's theorem). *Let G be a finite group, let F be a field of characteristic not dividing $|G|$, and let V be a representation of G over F. Then every subrepresentation of V is a direct summand of V as an $F[G]$-module.*

PROOF. Let W be an $F[G]$-submodule of V. As F-modules, we know that we can find a basis B' of W contained in a basis B of V. We then have a projection map $p\colon V \to W$ given by

$p(\sum_{v\in B} a_v v) = \sum_{w\in B'} a_w w$, where $a_v \in F$ equals zero for almost all v. This is an F-linear transformation that restricts to the identity on W, but it is not necessarily a $F[G]$-module homomorphism. So, define

$$\pi\colon V \to V, \qquad \pi(v) = \frac{1}{|G|}\sum_{g\in G} g^{-1}p(gv)$$

for $v \in V$. Then π is clearly F-linear, and moreover

$$\pi(hv) = \frac{1}{|G|}\sum_{g\in G} g^{-1}p(ghv) = \frac{1}{|G|}\sum_{k\in G} (kh^{-1})^{-1}p(kv) = hp(v),$$

so π is an $F[G]$-module homomorphism. Since W is an $F[G]$-submodule of V, the image of π is contained in W, and for $w \in W$, we have

$$\pi(w) = \frac{1}{|G|}\sum_{g\in G} g^{-1}p(gw) = \frac{1}{|G|}\sum_{g\in G} g^{-1}gw = w.$$

In particular, the inclusion of W in V splits π, so W is a direct summand of V as an $F[G]$-module. □

As a consequence of Maschke's theorem and Wedderburn theory, or more specifically, Theorem 13.1.8, we have the following corollary.

COROLLARY 13.3.2. *The group ring $F[G]$ is a semisimple F-algebra, which is to say isomorphic to a finite direct product of matrix rings over finite-dimensional division algebras over F.*

This in turn yields the following corollaries. For the first, see Corollary 13.1.24.

COROLLARY 13.3.3. *Let G be a finite abelian group, and let F be a field of characteristic not dividing $|G|$. Then $F[G]$ is a direct product of finite field extensions of F.*

EXAMPLE 13.3.4. By the Chinese remainder theorem, we have

$$\mathbb{Q}[\mathbb{Z}/p\mathbb{Z}] \cong \mathbb{Q}[x]/(x^p-1) \cong \mathbb{Q}[x]/(x-1) \times \mathbb{Q}[x]/\Phi_p(x) \cong \mathbb{Q} \times \mathbb{Q}(\zeta_p),$$

where ζ_p is a primitive pth root of unity in $\mathbb{C}$. Note, however, that if we take $\mathbb{F}_p$ in place of $\mathbb{Q}$, then we obtain

$$\mathbb{F}_p[\mathbb{Z}/p\mathbb{Z}] \cong \mathbb{F}_p[x]/(x^p-1) \cong \mathbb{F}_p[x]/(x-1)^p \cong \mathbb{F}_p[y]/(y^p)$$

for $y = x-1$, which is not a direct product of matrix rings over fields.

For the following, see Corollary 13.1.23.

COROLLARY 13.3.5. *If F is algebraically closed, then $F[G]$ is isomorphic to a direct product of matrix algebras over F.*

PROPOSITION 13.3.6. *Suppose that F is algebraically closed, and write*

$$F[G] \cong \prod_{i=1}^{k} M_{n_i}(F)$$

for some $k \geq 1$ and $n_i \geq 1$ for $1 \leq i \leq k$. Then k is equal to the number of conjugacy classes of $|G|$.

PROOF. First, we remark that $Z(M_{n_i}(F)) = F$, so $\dim_F Z(F[G]) = k$. For any $g \in G$, we form out of its conjugacy class C_g the sum $N_g = \sum_{h \in C_g} h$. If we let G act on $F[G]$ by conjugation, then the G-invariant module for this action is $Z(F[G])$. Moreover, N_g lies in this invariant group. That is, the action restricts to an action on G which preserves conjugacy classes, so

$$kN_gk^{-1} = \sum_{h \in C_g} khk^{-1} = \sum_{h \in C_g} h = N_g,$$

The elements N_g, where g runs over a set S of representatives for the conjugacy classes of G, are linearly independent as they are sums over disjoint sets of group elements. And if $z = \sum_{g \in G} a_g g \in Z(F[G])$ and $k \in G$, then

$$\sum_{g \in G} a_g g = \sum_{g \in G} a_g kgk^{-1} = \sum_{g \in G} a_{k^{-1}gk} g,$$

so $a_g = a_{k^{-1}gk}$ for all k, so $a_h = a_g$ for all $h \in C_g$. Thus, z is in the F-span of the elements N_g. Thus, we have $k = |S|$, the number of conjugacy classes. □

REMARK 13.3.7. If F is algebraically closed, then we may by Corollary 13.3.5 write

$$F[G] \cong \prod_{i=1}^{k} M_{n_i}(F)$$

for some $k \geq 1$ and $n_i \geq 1$ for $1 \leq i \leq k$. Then G has k isomorphism classes of irreducible representations of dimensions $n_1, n_2, \dots, n_k$. Let V_i be the ith of these, with $\dim_F V_i = n_i$. Then V_i occurs with multiplicity n_i in the regular representation $R[G]$, which is to say that $R[G] \cong V_1^{n_1} \oplus V_2^{n_2} \oplus \cdots \oplus V_k^{n_k}$. Under this isomorphism, each copy of V_i is identified with one of the simple left ideals in $M_{n_i}(F)$, isomorphic to the module F^{n_i} of column vectors for this ring. Counting dimensions tells us that

$$\sum_{i=1}^{k} n_i^2 = |G|.$$

EXAMPLE 13.3.8. The group S_3 has 3 conjugacy classes, so there are 3 isomorphism classes of irreducible representations of G, and the sum of the squares of their dimensions are 6, so they have dimensions 1, 1, and 2. Thus, we have

$$\mathbb{C}[S_3] \cong \mathbb{C} \times \mathbb{C} \times M_2(\mathbb{C}).$$

The two one-dimensional representations correspond to homomorphisms $G \to \mathbb{C}^\times$, factoring through $G^{\text{ab}} \cong \mathbb{Z}/2\mathbb{Z}$. There are exactly two of these, the trivial homomorphism and the sign map $\text{sign}\colon S_3 \to \{\pm 1\}$. These correspond to the trivial $F[G]$-module F and the $F[G]$-module F on which $\sigma \in S_3$ acts by $\sigma \cdot v = \text{sign}(\sigma)v$ for $v \in F$.

The irreducible two-dimensional representation W of S_3 is a subrepresentation of the 3-dimensional permutation representation $\rho_V\colon S_3 \to \text{GL}_3(\mathbb{C})$. That is, consider the standard basis $\{e_1, e_2, e_3\}$ of the corresponding $\mathbb{F}[S_3]$-module $V = \mathbb{F}^3$ on which S_3 acts by permuting the indices of the basis elements. Then W is spanned by $e_1 - e_2$ and $e_2 - e_3$. With respect to this basis, the corresponding homomorphism $\rho_W \colon S_3 \to \mathrm{GL}_2(\mathbb{C})$ satisfies

$$\rho_W((1\ 2)) = \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \rho_W((1\ 2\ 3)) = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}.$$

EXAMPLE 13.3.9. Since all of the $\mathbb{C}$-representations of S_3 take values in $\mathrm{GL}_n(\mathbb{Q})$ for some n, we have $\mathbb{Q}[S_3] \cong \mathbb{Q} \times \mathbb{Q} \times M_2(\mathbb{Q})$. In other words, the irreducible $\mathbb{C}$-representations of S_3 are obtained from the irreducible $\mathbb{Q}$-representations of S_3 by base change.

13.4. Characters

Recall from Lemma 9.6.22 that the traces of similar matrices are equal.

DEFINITION 13.4.1. Let V be a finite-dimensional vector space over a field F. The *trace* of $\varphi \in \mathrm{Aut}_F(V)$ is the trace of the matrix representing φ with respect to any choice of ordered basis of V.

DEFINITION 13.4.2. The *character* of a representation $\rho \colon G \to \mathrm{Aut}_F(V)$ of a group G on a finite-dimensional vector space V over a field F is a map $\chi \colon G \to F$ defined by

$$\chi(g) = \operatorname{tr} \rho(g).$$

TERMINOLOGY 13.4.3. We say that χ is a character of G if it is the character of a representation of G.

NOTATION 13.4.4. Given an $F[G]$-module V, we denote the corresponding representation (i.e., homomorphism) by ρ_V and and its character by χ_V.

EXAMPLES 13.4.5.

a. The character $\chi \colon G \to F$ of a one-dimensional representation $\rho \colon G \to F^\times$ satisfies $\chi(g) = \rho(g)$ for all $g \in G$.

b. The character of the permutation representation $\rho \colon S_n \to \mathrm{GL}_n(F)$ satisfies $\rho(\sigma) = |X_n^\sigma|$ for every $\sigma \in S_n$, where $X_n = \{1, 2, \ldots, n\}$.

c. Let W be as in Example 13.3.8. Then the character $\chi_W \colon S_3 \to \mathbb{C}$ satisfies $\chi_W((1\ 2)) = 0$ and $\chi_W((1\ 2\ 3)) = -1$.

d. The character χ of the regular representation $F[G]$ satisfies $\chi(1) = |G|$ and $\chi(g) = 0$ for all $g \in G - \{1\}$.

DEFINITION 13.4.6. The character of the trivial representation is called the *trivial character*, or *principal character* of G.

DEFINITION 13.4.7. A *character* $\chi_V \colon G \to F$ of an $F[G]$-module V is irreducible if V is irreducible.

DEFINITION 13.4.8. The degree of a character χ_V of an F-representation V of G is $\dim_F(V)$.

DEFINITION 13.4.9. A *class function* of G is a function $G \to F$, for F a field, that is constant on conjugacy classes in G.

LEMMA 13.4.10. *Let G be a group, let F be a field, and let V and W be F-representations of G. Then*

a. $\chi_V(e) = \dim_F V$,

b. $\chi_{V\oplus W} = \chi_V + \chi_W$,

c. $\chi_V = \chi_W$ *if V and W are isomorphic representations, and*

d. χ_V *is a class function on G.*

PROOF. Since $\rho_V(e)$ is the identity transformation, we have part *a*. Part b follows by choosing a basis of $V \oplus W$ that is a union of bases of V and W and noting that the matrix representing $\rho_{V\oplus W}(g)$ for $g \in G$ with respect to that basis is block diagonal with blocks $\rho_V(g)$ and $\rho_W(g)$. Part c holds as $\rho_V(g)$ and $\rho_W(g)$ are represented by similar matrices if V and W are isomorphic. Part d also holds as $\rho_V(g)$ and $\rho_V(g')$ are represented by similar matrices if g and g' are conjugate in G. □

PROPOSITION 13.4.11. *Let G be a finite group, and let F be a field of characteristic zero. Let V and W be finite-dimensional F-representations of G. Then V and W are isomorphic if and only if $\chi_V = \chi_W$.*

PROOF. By Lemma 13.4.10c, we know that $V \cong W$ implies $\chi_V = \chi_W$. Write

$$F[G] = \prod_{i=1}^{r} M_n(D_i).$$

For $1 \le i \le r$, let e_i denote the identity of $M_{n_i}(D_i)$, let V_i be the irreducible F-representation $D_i^{n_i}$ of G, and let χ_i denote its character. Then there exist m_i for $1 \le i \le r$ such that

$$V = \bigoplus_{i=1}^{r} V_i^{m_i}.$$

Extend χ_V by F-linearity to a map $\chi_V : F[G] \to F$. Then

$$\chi_V(e_j) = \sum_{i=1}^{r} m_i \chi_i(e_j) = m_i \dim_F V_i,$$

so the multiplicities m_i of the V_i in V are uniquely determined by χ_V. That is, χ_V determines the isomorphism class of V. □

The following easy lemma, which we will use implicitly, is also quite useful for passing between groups.

LEMMA 13.4.12. *Let V be an F-representation of G.*

a. If H is a subgroup of G, then V may be considered as an F-representation of H, and its character is the restriction $\chi_V|_H$.

b. If N is a normal subgroup of G and N acts trivially on V, then for $\pi\colon G \to G/N$ the quotient map, the character of V as an F-representation of G/N is $\pi \circ \psi_V$.

For the remainder of this section, we suppose that G is finite and F is algebraically closed of characteristic not dividing $|G|$.

PROPOSITION 13.4.13. *Suppose that G is finite and F is algebraically closed. Let V be a finite-dimensional F-representation of G. Then $\rho_V(g)$ is diagonalizable.*

PROOF. By restricting ρ_V to the cyclic subgroup generated by G, we may suppose that G is cyclic, say of order n. In this case, $F[G]$ is a direct sum of 1-dimensional representations V_i on which g acts by multiplication by ζ_n^i for ζ_n a choice of primitive nth root of unity in F. As V is semisimple, this tells us that V is a direct sum of 1-dimensional representations. The automorphism $\rho_V(g)$ is then diagonal with respect to any basis of V consisting of one basis element of each of these summands. □

Since g has finite order dividing $|G|$, the following corollary is immediate.

COROLLARY 13.4.14. *Let V be a finite-dimensional F-representation of G. Then the eigenvalues of $\rho_V(g)$ for $g \in G$ are all roots of unity of order dividing $|G|$.*

LEMMA 13.4.15. *Let V and W be finite-dimensional F-representations of G. Set $\overline{\chi_V}(g) = \chi_V(g^{-1})$ for all $g \in G$. Then we have*

a. $\chi_{V\otimes_F W} = \chi_V \chi_W$ and

b. $\chi_{\mathrm{Hom}_F(V,W)} = \overline{\chi_V}\chi_W$.

PROOF. By the commutativity of the tensor product and direct sums and the semisimplicity of $F[G]$, part a reduces to the case that V and W are irreducible. Through a simple application of Lemma 13.2.23, we may assume that F is algebraically closed. By Proposition 13.4.13, we may then diagonalize the matrices $\rho_V(g)$ and $\rho_W(g)$ for $g \in G$ with respect to choices of ordered bases $(v_1, \dots, v_n)$ of V and $(w_1, \dots, w_n)$ of W. We then have that $\rho_{V\otimes_F W}$ is diagonal with respect to the basis of elements $v_i \otimes w_j$ with respect to the lexicographical ordering. The diagonal coordinate corresponding to $v_i \otimes w_j$ is the product of the (i,i)-entry of $\rho_V(g)$ and the (j,j)-entry of $\rho_W(g)$. That is,

$$\operatorname{tr}\rho_{V\otimes W}(g) = (\operatorname{tr}\rho_V(g)) \cdot (\operatorname{tr}\rho_W(g)),$$

as desired.

For part b, we recall the isomorphism

$$\mathrm{Hom}_F(V,W) \cong V^* \otimes_F W$$

of Lemma 13.2.14. We are then reduced by part a to the case that $W = F$, the trivial $F[G]$-module. Again replacing F by its algebraic closure, we may diagonalize $\rho_V(g)$ for $g \in G$ with respect to a basis B of V. Let B^* be its dual basis. For $\phi \in B^*$ and $v \in B$, we have $\phi(g^{-1}v) = \alpha_v \phi(v)$, where $g^{-1}v = \alpha_v v$. Thus, the trace of $\rho_{V^*}(g)$ agrees with the trace of $\rho_V(g^{-1})$, as desired. □

REMARK 13.4.16. Let G be a finite group and F be a field of characteristic not dividing $|G|$. Since $\chi_{V\oplus W} = \chi_V + \chi_W$ and $\chi_{V\otimes_F W} = \chi_V \cdot \chi_W$, the set of F-valued characters of G form a ring with identity the trivial character.

PROPOSITION 13.4.17. *The irreducible F-characters of G form a basis for the F-vector space of F-valued class functions on G.*

PROOF. Let $g_1, \dots, g_r$ be representatives of the r conjugacy classes of G. The space of F-valued class functions of G has a basis consisting of the maps $\theta_i \colon G \to F$ for $1 \le i \le r$ such that $\theta_i(g) = 1$ if $g \in C_{g_i}$ and $\theta_i(g) = 0$ otherwise. On the other hand, there are also r irreducible F-representations V_i for $1 \le i \le r$ of G by Proposition 13.3.6, so it suffices to see that their characters $\chi_i = \chi_{V_i}$ are linearly independent.

Write $F[G] \cong \prod_{i=1}^r M_{n_i}(F)$ in such a way that V_i is the isomorphic to the simple module F^{n_i} of $M_{n_i}(F)$. Let e_i denote the idempotent of $F[G]$ corresponding to the identity of $M_{n_i}(F)$. We may extend χ_i F-linearly to a map $\chi_i \colon F[G] \to F$. Then $\chi_i(x)$ for $x \in F[G]$ is the trace of the endomorphism of V_i defined by left multiplication by x. Since left multiplication by e_i on V_i (resp., V_j for $j \neq i$) is the identity map (resp., zero map), we have $\chi_i(e_i) = n_i$ (resp., $\chi_i(e_j) = 0$ for $j \neq i$). Given any linear combination $\phi = \sum_{i=1}^r a_i \chi_i$ with $a_i \in F$, we have $\phi(e_j) = a_j n_j$, so $\phi = 0$ if and only if $a_i = 0$ for all i. □

We can identify the idempotents in $F[G]$ that correspond to identity matrices in terms of characters.

PROPOSITION 13.4.18. *Let χ_i for $1 \le i \le r$ denote the irreducible F-characters of G, and let n_i denote the degree of χ_i Then the elements*

$$e_i = \frac{n_i}{|G|} \sum_{g \in G} \chi_i(g^{-1}) g$$

are the primitive, central, orthogonal idempotents of $F[G]$.

PROOF. Let f_i denote the primitive central idempotent in $F[G]$ that acts on the identity on the irreducible representation V_i with character χ_i. Write $f_i = \sum_{g \in G} a_g g$ with $a_g \in F$ for $g \in G$. For any $g \in G$, we have

$$\chi_{F[G]}(f_i g^{-1}) = \sum_{h \in G} \chi_{F[G]}(a_h h g^{-1}) = a_g |G|.$$

On the other hand, we have $\chi_{F[G]} = \sum_{i=1}^r n_i \chi_i$, where $n_i = \dim_F V_i$, so

$$\chi_{F[G]}(f_i g^{-1}) = \sum_{j=1}^r n_j \chi_j(f_i g^{-1}).$$

If $\rho_i \colon G \to \mathrm{End}_F(V_i)$ is the F-linear map restricting to ρ_{V_i}, then

$$\rho_j(f_i g^{-1}) = \rho_j(f_i)\rho_j(g^{-1}) = \delta_{ij}\rho_j(g^{-1}),$$

so $\chi_j(f_i g^{-1}) = \delta_{ij}\chi_j(g^{-1})$. Thus, we have

$$a_g |G| = n_i \chi_i(g^{-1}).$$

It follows that

$$f_i = \frac{n_i}{|G|} \sum_{g \in G} \chi_i(g^{-1}) g.$$

□

13.5. Character tables

In this section, we focus on the theory of $\mathbb{C}$-valued characters of a finite group G.

DEFINITION 13.5.1. A *character table* of a finite group G is a matrix in $M_r(\mathbb{C})$, where r is the number of conjugacy classes of G with (i,j)-entry $\chi_i(g_j)$, where χ_i for $1 \le i \le r$ are the distinct characters of the irreducible $\mathbb{C}$-representations of G and g_i for $1 \le i \le r$ are representatives of the distinct conjugacy classes in G.

Usually, a character table is written in a table format, as in the following example.

EXAMPLE 13.5.2. Take the group S_3. Let χ_1 denote the trivial character, χ_2 denote the sign character, and χ_3 the irreducible character of dimension 2. By Example 13.3.8, the character table of S_3 is then as follows:

S_3	1	$(1\ 2)$	$(1\ 2\ 3)$
χ_1	1	1	1
χ_2	1	-1	1
χ_3	2	0	-1

Recall that $\overline{\alpha}$ denotes the complex conjugate of a complex number α.

LEMMA 13.5.3. *Let χ be a $\mathbb{C}$-valued character of degree d of a finite group G of order n. For $g \in G$, we have $\chi(g) \in \mathbb{Z}[\zeta_n]$, $|\chi(g)| \le d$, and $\chi(g^{-1}) = \overline{\chi(g)}$.*

PROOF. Let $\rho\colon G \to \mathrm{Aut}_{\mathbb{C}}(V)$ be the representation corresponding to χ. By Corollary 13.4.14, the value $\rho(g)$ can be diagonalized to matrix with entries in μ_n. Since the inverse of a root of unity is its complex conjugate, $\rho(g^{-1}) = \rho(g)^{-1}$ may be then be represented by the diagonal matrix $A^{-1} = (\overline{a_{ij}})$. Then $\chi(g)$ is a sum of d roots of unity of order dividing n, which the first two statements, and we have

$$\chi(g^{-1}) = \mathrm{tr}(A^{-1}) = \overline{\mathrm{tr}(A)} = \overline{\chi(g)}.$$

□

Let us set $\zeta_n = e^{2\pi i/n} \in \mathbb{C}$ for $n \ge 1$. The following lemma is useful for producing new characters out of old.

LEMMA 13.5.4. *Let G be a group of order n, and let $\sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$, where $\zeta_n = e^{2\pi i/n} \in \mathbb{C}$. If χ is a character of G, then so is $\chi^\sigma\colon G \to F$ defined by $\chi^\sigma(g) = \sigma(\chi(g))$ for all $g \in G$. Moreover, if $a \in (\mathbb{Z}/n\mathbb{Z})^\times$ is such that $\sigma(\zeta_n) = \zeta_n^a$, then $\chi^\sigma(g) = \chi(g^a)$.*

PROOF. By Proposition 13.4.13, if G is finite of order n, then every $\mathbb{C}$-representation of G is the base change of a $\mathbb{Q}(\zeta_n)$-representation. Let V be the $\mathbb{Q}(\mu_n)$-representation of G with character χ. As a vector space, $V \cong \mathbb{Q}(\mu_n)^d$ for some $d \ge 0$, and so σ induces an automorphism $\sigma\colon V \to V$ as the direct sum of the automorphisms $\mathbb{Q}(\mu_n) \to \mathbb{Q}(\mu_n)$. Then $\sigma \circ \rho_V$ is again a representation, and its character is χ^σ. Note that $\sigma \circ \rho_V(g) = \rho_V(g^a)$, since in diagonalized form, the entries of $\rho_V(g)$ are all elements of μ_n upon which σ acts by raising to the ath power. □

We pause for a moment to discuss inner products on $\mathbb{C}$-vector spaces.

DEFINITION 13.5.5. An *inner product* on an $\mathbb{C}$-vector space V is a map $\langle\ ,\ \rangle\colon V\times V\to\mathbb{C}$ that satisifes

$$\langle \alpha v+v',w\rangle=\alpha\langle v,w\rangle+\langle v',w\rangle \quad\text{and}\quad \langle v,\beta w+w'\rangle=\bar{\beta}\langle v,w\rangle+\langle v,w'\rangle$$

for all $v,v',w,w'\in V$ and $\alpha,\beta\in\mathbb{C}$.

TERMINOLOGY 13.5.6. That $\langle\ ,\ \rangle$ is an inner product on a $\mathbb{C}$-vector space V may be expressed as saying that it is left $\mathbb{C}$-linear (or just linear) and right conjugate linear.

DEFINITION 13.5.7. An inner product $\langle\ ,\ \rangle$ on a $\mathbb{C}$-vector space V is *positive definite* if $\langle v,v\rangle\geq 0$ for all $v\in V$, with equality only for $v=0$.

DEFINITION 13.5.8. An inner product $\langle\ ,\ \rangle$ on a $\mathbb{C}$-vector space V is *Hermitian* if it is positive definite and $\langle v,w\rangle=\overline{\langle w,v\rangle}$ for all $v,w\in V$.

DEFINITION 13.5.9. A basis B of a $\mathbb{C}$-vector space V with a Hermitian inner product $\langle\ ,\ \rangle$ is *orthonormal* if $\langle v,w\rangle=\delta_{v,w}$ for all $v,w\in B$.

DEFINITION 13.5.10. A *complex inner product space* is a pair consisting of a $\mathbb{C}$-vector space V and a Hermitian inner product on V.

EXAMPLE 13.5.11. The dot product on $\mathbb{C}^n$ defined by

$$(a_1,a_2,\dots,a_n)\cdot(b_1,b_2,\dots,b_n)=\sum_{i=1}^{n}a_i\overline{b_i}$$

is a positive definite, Hermitian inner product on $\mathbb{C}^n$. The standard basis of $\mathbb{C}^n$ is orthonormal with respect to the dot product, so $\mathbb{C}^n$ is an inner product space with respect to the dot product.

DEFINITION 13.5.12. An inner product $\langle\ ,\ \rangle$ on a $\mathbb{C}$-representation V of G is said to be *G-invariant*, or an *invariant inner product*, if $\langle gv,gw\rangle=\langle v,w\rangle$ for all $v,w\in V$.

The following provides a useful example.

LEMMA 13.5.13. *Let $\langle\ ,\ \rangle$ be a Hermitian inner product on a $\mathbb{C}$-representation V of G. Then the map $[\ ,\ ]\colon V\times V\to\mathbb{C}$ defined on $v,w\in V$ by*

$$[v,w]=\frac{1}{|G|}\sum_{g\in G}\langle gv,gw\rangle$$

is a G-invariant inner product on V.

PROOF. As a positive real scalar multiple of a sum of Hermitian inner products on V, the pairing $[\ ,\ ]$ is also Hermitian. The invariance by an element of G follows by reindexing the sum. □

The next lemma contains the definition of an inner product on the space of $\mathbb{C}$-valued class functions of G.

LEMMA 13.5.14. *The function which assigns to a pair (θ, ψ) of $\mathbb{C}$-valued class function of G the value*

$$\langle \theta, \psi \rangle = \frac{1}{|G|} \sum_{g \in G} \theta(g) \overline{\psi(g)}$$

is a positive definite, Hermitian inner product on the space of $\mathbb{C}$-valued class functions of G.

We consider the space of class functions as a Hermitian inner product space with respect to the Hermitian inner product of Lemma 13.5.14.

REMARK 13.5.15. If we let $h \in G$ act on the space of class functions on G by $(h \cdot \theta)(g) = \theta(h^{-1}g)$, then the resulting inner product is G-invariant.

REMARK 13.5.16. The inner product of the characters of any two $\mathbb{C}$-representations V and W is real by Lemma 13.5.3, since

$$\sum_{g \in G} \chi_V(g) \overline{\chi_W(g)} = \sum_{g \in G} \chi_V(g^{-1}) \overline{\chi_W(g^{-1})} = \sum_{g \in G} \overline{\chi_V(g)} \chi_W(g).$$

LEMMA 13.5.17. *Let V be a $\mathbb{C}[G]$-module of finite dimension. Then*

$$\dim_{\mathbb{C}} V^G = \frac{1}{|G|} \sum_{g \in G} \chi_V(g).$$

PROOF. Let $z = \frac{1}{|G|} N_G \in \mathbb{C}[G]$. Since $N_G^2 = |G|$, the element z is an idempotent. The $\mathbb{C}$-linear endomorphism T of V defined by left multiplication by z therefore has minimal polynomial dividing $x^2 - x = x(x-1)$. In particular, it is diagonalizable. The trace of T is then the sum of its nontrivial eigenvalues, which is the dimension of the eigenspace $E_1(T)$ of 1. It remains then only to show that $E_1(T) = V^G$. We check this on $v \in V$: if $zv = v$, then $gv = gzv = zv = v$ for all $g \in G$, while if $gv = v$ for all $g \in G$, then $zv = \frac{1}{|G|} |G| v = v$. □

PROPOSITION 13.5.18. *Let V and W be complex G-representations. Then*

$$\langle \chi_V, \chi_W \rangle = \dim_{\mathbb{C}} \mathrm{Hom}_{\mathbb{C}[G]}(V, W).$$

PROOF. Note that $\mathrm{Hom}_{\mathbb{C}[G]}(V, W) = \mathrm{Hom}_{\mathbb{C}}(V, W)^G$, where $g \in G$ acts on $\phi \in \mathrm{Hom}_{\mathbb{C}}(V, W)$ by $(g \cdot \phi)(v) = g\phi(g^{-1}v)$ for every $v \in V$. Thus,

$$\dim_{\mathbb{C}} \mathrm{Hom}_{\mathbb{C}[G]}(V, W) = \frac{1}{|G|} \sum_{g \in G} \chi_{\mathrm{Hom}_{\mathbb{C}}(V,W)}(g) = \frac{1}{|G|} \sum_{g \in G} \overline{\chi_V(g)} \chi_W(g),$$

the last step by Lemma 13.4.15b. □

We may now prove the orthogonality of the basis of characters.

THEOREM 13.5.19 (First orthogonality relation). *The set of irreducible complex characters of a finite group G forms an orthonormal basis of the space of $\mathbb{C}$-valued class functions of G.*

PROOF. Let V_i for $1 \leq i \leq r$ be the distinct irreducible $\mathbb{C}[G]$-modules, and let χ_i denote the character of V_i. For any $1 \leq i, j \leq r$, we have

$$\langle \chi_i, \chi_j \rangle = \dim_{\mathbb{C}} \operatorname{Hom}_{\mathbb{C}[G]}(V_i, V_j),$$

by Proposition 13.5.18. The result then follows by Schur's lemma. □

This orthogonality also gives us a sort of orthogonality of rows of the character table.

REMARK 13.5.20. Let $\{g_1, g_2, \ldots, g_r\}$ be a set of representatives of the conjugacy classes of a finite group G. Let $\chi_1, \chi_2, \ldots, \chi_r$ be the complex irreducible character. By Theorem 13.5.19, the rows r_i of the character table with (i, j)-entry $\chi_i(g_j)$ are orthogonal with respect to the weighted dot product

$$r_i \cdot r_{i'} = \frac{1}{|G|} \sum_{j=1}^{r} c_j \chi_i(g_j) \overline{\chi_{i'}(g_j)} = \langle \chi_i, \chi_{i'} \rangle,$$

where the weight c_j is the order of the conjugacy class of g_j.

We also have an orthogonality relation for columns.

THEOREM 13.5.21 (Second orthogonality relation). *Let G be a finite group, and let $\chi_1, \chi_2, \ldots, \chi_r$ be its distinct irreducible, complex characters. For any $g, h \in G$, we have*

$$\sum_{i=1}^{r} \chi_i(g) \overline{\chi_i(h)} = \begin{cases} |Z_g| & \text{if } g \text{ and } h \text{ are conjugate}, \\ 0 & \text{otherwise}, \end{cases}$$

where Z_g denotes the centralizer of g in G.

PROOF. Let $g_1, g_2, \ldots, g_r$ represent the distinct conjugacy classes in G, and let $A \in M_r(\mathbb{C})$ be the matrix with (i, j)-entry $\chi_i(g_j)$. Let C be the diagonal matrix with (i, i)-entry $c_i = |C_{g_i}|$. Then

$$(AC\bar{A}^t)_{ij} = \sum_{k=1}^{r} \chi_i(g_k) c_k \overline{\chi_j(g_k)} = \delta_{ij} |G|,$$

the last step by Remark 13.5.20. In particular, $AC\bar{A}^t$ is a scalar multiple of the identity matrix, so $AC\bar{A}^t = \bar{A}^t AC$, which tells us that

$$\delta_{ij} |G| = (\bar{A}^t AC)_{ij} = \sum_{k=1}^{r} \overline{\chi_k(g_i)} \chi_k(g_j) c_j.$$

Since $|Z_{g_j}| = |G| c_j^{-1}$ by the orbit-stabilizer theorem, we are done. □

Let us study character tables in some examples.

EXAMPLE 13.5.22. Let $n \geq 1$, and let $\zeta = e^{2\pi i/n} \in \mathbb{C}$. Every character of $\mathbb{Z}/n\mathbb{Z}$ is a power of the character $\chi \colon \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}^\times$ given by $\chi(i) = \zeta^i$. Since $\mathbb{Z}/n\mathbb{Z}$ is abelian, every element of $\mathbb{Z}/n\mathbb{Z}$ is the lone element in its conjugacy class. The character table of $\mathbb{Z}/n\mathbb{Z}$ is as follows:

$\mathbb{Z}/n\mathbb{Z}$	0	1	2	$\cdots$	$n-1$
1	1	1	1	$\cdots$	1
χ	1	ζ	ζ^2	$\cdots$	ζ^{n-1}
χ^2	1	ζ^2	ζ^4	$\cdots$	ζ^{n-2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
χ^{n-1}	1	ζ^{n-1}	ζ^{n-2}	$\cdots$	$\zeta.$

REMARK 13.5.23. In general, the number of 1-dimensional complex characters of a finite group G is $|G^{\text{ab}}|$, since these are exactly the irreducible representations of $\mathbb{C}[G^{\text{ab}}]$, which has $\mathbb{C}$-dimension $|G^{\text{ab}}|$.

EXAMPLE 13.5.24. Let G be any nonabelian group of order 8. By Theorem 7.5.2, there are two up to isomoprhism, D_4 and Q_8. The center Z of G has order 2 (it is nontrivial since G is a 2-group and if it had order at least 4, it is easy to see that the group would be abelian). Furthermore, G/Z is abelian since all groups of order 4 are abelian. This also means that G/Z is the abelianization of G. It is also easy to see that this implies $G/Z \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$. So G has four characters $\chi_1, \cdots, \chi_4$ of degree 1 and therefore one character χ_5 of degree 2 to make $8 = 2^2 + 1 + 1 + 1 + 1$. Pick representatives g and h in G of the two summands $\mathbb{Z}/2\mathbb{Z}$. Then g, h, and gh must be representatives of distinct conjugacy classes, which are then forced to have order 2 since $g, h, gh \notin Z$. Finally, let z generate the center. The character table is

G	1	z	g	h	gh
χ_1	1	1	1	1	1
χ_2	1	1	1	-1	-1
χ_3	1	1	-1	1	-1
χ_4	1	1	-1	-1	1
χ_5	2	-2	0	0	0

The last row is determined by orthogonality of the columns, since its first entry must be 2. Note that this implies that the isomorphism type of a group is not determined by its character.

EXAMPLE 13.5.25. Note that S_4 has 5 conjugacy classes, and the sums of the squares of the degrees n_i of the 5 irreducible characters χ_i equals $|24|$. Also $S_4 \to \mathbb{Z}/2\mathbb{Z}$ via the sign map, so there are (at least) two 1-dimensional characters: the trivial character χ_1 and the sign character χ_2. Since $n_3^2 + n_4^2 + n_5^2 = 22$, we have $n_3 = 2$ and $n_4 = n_5 = 3$ (if put in increasing order). The quotient of S_4 by the normal subgroup $\langle (1\ 2)(3\ 4), (1\ 3)(2\ 4) \rangle$ is isomorphic to S_3, so we obtain by composition with the irreducible two-dimensional representation $\rho \colon S_3 \to \mathrm{GL}_2(\mathbb{C})$ a two-dimensional representation $\chi_3 \colon S_4 \to \mathrm{GL}_2(\mathbb{C})$, which is nonabelian and hence irreducible, being semisimple. Whatever χ_4 is, note that if we tensor its representation V_4 with the representation V_2 of the sign character χ_2, we obtain another irreducible character $\chi_{V_4 \otimes_{\mathbb{C}} V_2} = \chi_4 \chi_2$ of dimension 3, which we call χ_5. (We will see that it is actually different from χ_4.) The character table is

S_4	e	$(1\ 2)$	$(1\ 2\ 3)$	$(1\ 2\ 3\ 4)$	$(1\ 2)(3\ 4)$
χ_1	1	1	1	1	1
χ_2	1	-1	1	-1	1
χ_3	2	0	-1	0	2
χ_4	3	1	0	-1	-1
χ_5	3	-1	0	1	-1.

The entries in blue are determined from the character table for S_3. The entries in red are obtained by noting that $\chi_5 = \chi_4\chi_2$ and using orthogonality of columns. (That is, the third and fifth columns are determined using orthogonality with the first and the second and fourth, up to sign, using orthogonality with the first and each other.)

Note that we can restrict representations and characters to subgroups: for $H \leqslant G$ and a G-representation V, this amounts to considering V as a module over the group ring of H and restricting the function χ_V to H.

EXAMPLE 13.5.26. The group A_4 has 4 conjugacy classes with representatives e, $(1\ 2\ 3)$, $(1\ 3\ 2)$, and $(1\ 2\ 3\ 4)$. We have $A_4^{\mathrm{ab}} \cong \mathbb{Z}/3\mathbb{Z}$, generated by the image of $(1\ 2\ 3)$, so there are three abelian characters which are the powers of the character χ such that $\chi((1\ 2\ 3)) = \omega$, where $\omega = e^{2\pi i/3}$. Since $|A_4| = 12$, there is one more character ψ, which has degree 3. Its values can be calculated by orthogonality of columns, yielding the character table

A_4	e	$(1\ 2\ 3)$	$(1\ 3\ 2)$	$(1\ 2)(3\ 4)$
1	1	1	1	1
χ	1	ω	ω^2	1
χ^2	1	ω^2	ω	1
ψ	3	0	0	-1.

13.6. Induced representations

Let G be a finite group and H a subgroup. For a commutative ring R we can view a $R[G]$-module A as an $R[H]$-module in the obvious fashion. When thinking of A as an $R[H]$-module, it is often helpful to give it a new name and symbol

DEFINITION 13.6.1. An $R[G]$-module A viewed as an $R[H]$-module is called the *restriction* of A from G to H and is denoted by $\mathrm{Res}^G_H(A)$.

Together with the obvious definition on morphisms, restriction defines an exact functor

$$\mathrm{Res}^G_H \colon R[G]\textbf{-mod} \to R[H]\textbf{-mod}.$$

The natural question arises as to whether or not Res^G_H has an adjoint, and indeed, it has a left adjoint. We first give the construction.

DEFINITION 13.6.2. Let H be a subgroup of a group G, and let R be a commutative ring. The *induced module* from H to G of an $R[H]$-module B is the $R[G]$-module

$$\mathrm{Ind}^G_H(B) = \mathrm{Hom}_{R[H]}(R[G], B)$$

where for $\varphi \in \mathrm{Ind}^G_H(B)$, we let $g \in G$ act by $(g \cdot \varphi)(x) = \varphi(xg)$ for all $x \in F[G]$.

REMARK 13.6.3. Since $R[G]$ is $R[H]$-free, Ind_H^G provides an exact functor from $R[H]$-**mod** to $R[G]$-**mod**, since is the functor $h_{R[G]}$ in our earlier notation.

If H is of finite index in G, we have an alternate description of the induced module. That is, there is another way in which to produce an $R[G]$-module from an $R[H]$-module B using tensor products. That is, we can take the $R[G]$-module $R[G] \otimes_{R[H]} B$, where $g \in G$ acts on $x \otimes b$ in the tensor product by $g(x \otimes b) = gx \otimes b$.

PROPOSITION 13.6.4. *Let H be a finite index subgroup of G, and let R be a commutative ring. Given an $R[H]$-module B, there is natural isomorphism*

$$\kappa \colon \mathrm{Ind}_H^G(B) \xrightarrow{\sim} R[G] \otimes_{R[H]} B$$

given on $\varphi \in \mathrm{Ind}_H^G(B)$ by

$$\kappa(\varphi) = \sum_{\bar{g} \in H \backslash G} g^{-1} \otimes \varphi(g),$$

where for each $\bar{g} \in H\backslash G$, the element $g \in G$ is a choice of representative of $\bar{g}$.

PROOF. First, we note that χ is a well-defined map, as

$$(hg)^{-1} \otimes \varphi(hg) = g^{-1}h^{-1} \otimes h\varphi(g) = g \otimes \varphi(g)$$

for $\varphi \in R[G] \otimes_{R[H]} B$, $h \in H$, and $g \in G$. Next, we see that χ is an $R[G]$-module homomorphism, as

$$\chi(g'\varphi) = \sum_{\bar{g} \in H \backslash G} g^{-1} \otimes \varphi(gg') = g' \sum_{\bar{g} \in H \backslash G} (gg')^{-1} \otimes \varphi(gg') = g'\chi(\varphi)$$

for $g' \in G$. As the coset representatives form a basis of $R[G]$ as a free $R[H]$-module, we may define an inverse to χ that maps

$$\sum_{\bar{g} \in H \backslash G} g^{-1} \otimes b_g \in \mathrm{Ind}_H^G(B)$$

to the unique $R[H]$-linear map φ that takes the value b_g on g for the chosen representative of $\bar{g} \in H\backslash G$. □

PROPOSITION 13.6.5. *Let H be a finite index subgroup of G. Then* Ind_H^G *is left adjoint to* Res_H^G.

PROOF. Using the alternate characterization of $\mathrm{Ind}_H^G U$ of Proposition 13.6.4 and the adjointness of Hom and $\otimes$, we have

$$\mathrm{Hom}_{R[G]}(R[G] \otimes_{R[H]} U, V) \cong \mathrm{Hom}_{R[H]}(U, \mathrm{Hom}_{R[G]}(R[G], V)) \cong \mathrm{Hom}_{R[H]}(U, V),$$

the latter isomorphism being induced by evaluation at 1 in the second variable. □

DEFINITION 13.6.6. Let H be a subgroup of a group G, let W be an F-representation of H.

a. The *induced representation* from H to G of W is $\mathrm{Ind}_H^G(W)$.

b. If H has finite index in G and W is finite-dimensional with character ψ, then the *induced character* $\mathrm{Ind}_H^G(\psi)$ of ψ is the character of $\mathrm{Ind}_H^G(W)$.

EXAMPLE 13.6.7. Let H be a finite index subgroup of G. For the trivial representation F of H, we have

$$\operatorname{Ind}_H^G(F) \cong F[G] \otimes_{F[H]} F \cong F[G/H],$$

where the latter module is the $F[G]$-module with F-basis the left G-set G/H. That is, $\operatorname{Ind}_H^G(F)$ is the permutation representation for the left action of G on G/H. Thus, the induced character $\chi = \operatorname{Ind}_H^G(\psi_1)$ of the trivial character ψ_1 on H has the property that $\chi(g)$ is the number of left H-cosets fixed by left multiplication by $g \in G$.

REMARK 13.6.8. Let H be a finite index subgroup of G. The induced representation of the regular representation $F[H]$ of H is the regular representation $F[G]$ of G. In particular, all irreducible F-representations of G are summands of induced representations of the irreducible F-representations of H.

REMARK 13.6.9. In finite group theory, one often uses the alternate tensor product characterization provided by Proposition 13.6.4 as the definition of the induced representation.

NOTATION 13.6.10. For a subgroup H of a group G, we denote the restriction of a character χ of G to H by $\operatorname{Res}_H^G \chi$, or more simply $\chi|_H$.

NOTATION 13.6.11. For the inner product of Lemma 13.5.14, we use $\langle\ ,\ \rangle_G$ to indicate its dependence on the group G.

DEFINITION 13.6.12. For a character χ of a finite-dimensional $\mathbb{C}$-representation of a group, the *multiplicity* of an irreducible character ψ in it is the multiplicity of the irreducible representation with character ψ in the representation with character χ.

We have the following corollary of Proposition 13.6.4.

COROLLARY 13.6.13 (Frobenius reciprocity). *Let G be a finite group and H a subgroup. Let ψ be a $\mathbb{C}$-valued character of H and χ be a $\mathbb{C}$-valued character of G. Then*

$$\langle \operatorname{Ind}_H^G \psi, \chi \rangle_G = \langle \psi, \operatorname{Res}_H^G \chi \rangle_H.$$

PROOF. Let $\psi = \chi_U$ and $\chi = \chi_V$ for representations U and V of H and G, respectively. By Propositions 13.5.18 and 13.6.5, we have

$$\langle \operatorname{Ind}_H^G \psi, \chi \rangle_G = \dim_{\mathbb{C}} \operatorname{Hom}_{\mathbb{C}[G]}(\operatorname{Ind}_H^G U, V) = \dim_{\mathbb{C}} \operatorname{Hom}_{\mathbb{C}[H]}(U, V) = \langle \psi, \operatorname{Res}_H^G \chi \rangle_H.$$

□

We can construct tables that contain these values of the pairings in Corollary 13.6.13.

DEFINITION 13.6.14. The *induction-restriction table* of a subgroup H of a finite group G is the matrix with rows indexed by the complex irreducible characters ψ_i of H and columns by the complex irreducible characters χ_j of G with (i,j)-entry $\langle \psi_i, \operatorname{Res}_H^G \chi_j \rangle_H$.

EXAMPLE 13.6.15. Again let χ_i for $1 \le i \le 5$ and 1, χ, χ^2, and ψ be the characters of S_4 and A_4, respectively of Examples 13.5.25 and 13.5.26. From the character tables of $G = S_4$ and $H = A_4$, we see that $\chi_1|_H = \chi_2|_H = 1$, $\chi_3|_H = \chi + \chi^2$, and $\chi_4|H = \chi_5|_H = \psi$. The induction-restriction table is

	χ_1	χ_2	χ_3	χ_4	χ_5
1	1	1	0	0	0
χ	0	0	1	0	0
χ^2	0	0	0	1	0
ψ	0	0	0	1	1.

Frobenius reciprocity tells us that $\mathrm{Ind}_H^G 1 = \chi_1 + \chi_2$, $\mathrm{Ind}_H^G \chi = \mathrm{Ind}_H^G \chi^2 = \chi_3$, and $\mathrm{Ind}_H^G \psi = \chi_4 + \chi_5$.

PROPOSITION 13.6.16. *Let H be a finite index subfroup of a group G, and let $g_1, \dots, g_k$ be a system of left coset representatives of H in G. For a character ψ of H, extend ψ to a function $\tilde{\psi} \colon G \to \mathbb{C}$ by setting $\tilde{\psi}(g) = 0$ if $g \notin H$. For $g \in G$, we then have*

$$\mathrm{Ind}_H^G(\psi)(g) = \sum_{i=1}^{k} \tilde{\psi}(g_i^{-1} g g_i).$$

PROOF. Let W be an m-dimensional representation of H with character ψ, and let $B = (w_1, \dots, w_m)$ be an ordered F-basis of W. Recall that

$$\mathrm{Ind}_H^G(W) \cong F[G] \otimes_{F[H]} W$$

We have a basis $g_i \otimes w_j$ of $F[G] \otimes_{F[H]} W$ for $1 \le i \le k$ and $1 \le j \le m$ with the lexicographical ordering.

For a given $1 \le i \le k$, any $g \in G$ satisfies

$$g g_i = g_{\sigma(i)} h_i$$

for some $\sigma \in S_k$ and $h_i \in H$. Then

$$g(g_i \otimes w_j) = g_{\sigma(i)} \otimes h_i w_j.$$

With respect to the given basis, the matrix of $\rho_W(g)$ is a k-by-k matrix of blocks in $M_m(F)$ with one nonzero block in each row and each column, i.e., the blocks with coordinates $(i, \sigma(i))$, which are those representing $\rho_W(h_i)$ with respect to the basis B.

Adding up the diagonal entries in the (i,i)-block, we get 0 if $i' \neq i$ and $\psi(h) = \psi(g_i^{-1} g g_i)$ if $i' = i$. By definition of $\tilde{\psi}$, this equals $\tilde{\psi}(g_i^{-1} g g_i)$ in all cases. Summing over i, we obtain the result. □

COROLLARY 13.6.17. *Let G be a finite group and H be a subgroup of G. Let ψ be a character of H. For $g \in G$, we then have*

$$\mathrm{Ind}_H^G(\psi)(g) = \frac{1}{|H|} \sum_{\substack{k \in G \\ kgk^{-1} \in H}} \psi(k^{-1} g k).$$

PROOF. This follows from the formula of Proposition 13.6.16. To see that, take $\tilde{\psi}$ as in its statement, and note that for any $h \in H$, we have

$$\tilde{\psi}((g_i h)^{-1} g (g_i h)) = \tilde{\psi}(h^{-1}(g_i^{-1} g g_i) h),$$

since conjugation by h^{-1} preserves H and $G - H$, and ψ is a class function on H. □

The following corollary is immediate.

COROLLARY 13.6.18. *Let H be a finite index normal subgroup of G, and let χ be a character of H. Then* $\mathrm{Ind}_H^G(\psi)(g) = 0$ *for all $g \notin H$.*

EXAMPLE 13.6.19. Consider the dihedral group $G = D_{2n}$ of order $4n$ with $n \geq 3$. It has abelianization the Klein 4-group generated by the images of r and s, so G has four degree 1 characters $\chi_1, \chi_2, \chi_3, \chi_4$ with χ_1 trivial, $\chi_2(r) = \chi_3(s) = -1$ and $\chi_2(s) = \chi_3(r) = 1$, and $\chi_4 = \chi_2\chi_3$. Now, consider the cyclic subgroup $H = \langle r \rangle$, which has characters ψ^i for $0 \leq i \leq 2n-1$ with $\psi(r) = \zeta_{2n}$. The induced character θ_i of ψ^i is trivial on all reflections and satisfies

$$\theta_i(r^j) = \zeta_{2n}^{ij} + \zeta_{2n}^{-ij}.$$

The characters $\theta_i = \mathrm{Ind}_H^G \psi^i$ with $1 \leq i \leq n-1$ are all distinct degree 2 characters which are clearly not sums of the χ_i, so they are irreducible characters of G. The sum of the squares of the dimensions of these characters is $4 \cdot 1^2 + (n-1) \cdot 2^2 = 4n = |G|$, so these are all of the irreducible characters on G. Setting

$$\xi_k = \zeta_{2n}^k + \zeta_{2n}^{-k} = 2\cos(k\pi/n) \in \mathbb{R},$$

the character table is then as follows.

D_{2n}	e	s	rs	r	r^2	$\cdots$	r^{n-1}	r^n
χ_1	1	1	1	1	1	$\cdots$	1	1
χ_2	1	1	-1	-1	1	$\cdots$	$(-1)^{n-1}$	$(-1)^n$
χ_3	1	-1	-1	1	1	$\cdots$	1	1
χ_4	1	-1	1	-1	1	$\cdots$	$(-1)^{n-1}$	$(-1)^n$
θ_1	2	0	0	ξ_1	ξ_2	$\cdots$	ξ_{n-1}	-2
θ_2	2	0	0	ξ_2	ξ_4	$\cdots$	$\xi_{2(n-1)}$	2
θ_3	2	0	0	ξ_3	ξ_6	$\cdots$	$\xi_{3(n-1)}$	-2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
θ_{n-1}	2	0	0	ξ_{n-1}	$\xi_{2(n-1)}$	$\cdots$	$\xi_{(n-1)^2}$	$(-1)^{n-1}2.$

Here, one might note that $\xi_0 = 2$, and $\xi_k = \xi_{k+2n} = -\xi_{k+n} = -\xi_{n-k}$ for all k. We remark that $\theta_i|_H = \psi^i + \psi^{-i}$ for all $1 \leq i \leq n-1$, while $\theta_0 = \chi_1 + \chi_3$ and $\theta_n = \chi_2 + \chi_4$, consistent with Frobenius reciprocity.

We also give a formula which tells us explicitly how to determine the induced character to G of an H-character from the character table for H and knowledge of conjugacy classes.

PROPOSITION 13.6.20. *Let G be a finite group and H be a subgroup of G. Let ψ be a character of G, let $g \in G$, and let C_g be the conjugacy class of g in G. Write $H \cap C_g$ as a possibly empty disjoint union of conjugacy classes $T_1, \ldots, T_l$ of H. For $1 \leq i \leq l$, let h_i be a representative of T_i. Then*

$$\mathrm{Ind}_H^G \psi(g) = [G:H] \sum_{i=1}^{l} \frac{|T_i|}{|C_g|} \psi(h_i).$$

PROOF. This is a matter of counting. That is, by Corollary 13.6.17, we must show that the number of $k \in G$ such that $k^{-1}gk$ is conjugate to h_i in H is $|Z_g||T_i|$, where Z_g denotes the centralizer of g in G. We know that there are $|Z_g|$ elements of G that conjugate g to any particular element of C_g. Thus, there are $|Z_g||T_i|$ elements in G that conjugate g to one of the elements in T_i, as desired. □

We may use Proposition 13.6.20 to determine the induced characters on a group from the characters on its subgroup.

EXAMPLE 13.6.21. Take $H = S_3$, which we view as a subgroup of $G = S_4$. Recall that the conjugacy classes of S_4 are determined by cycle type, with conjugacy classes $C_1, \cdots, C_5$ corresponding to cycle types e, (1 2), (1 2 3), (1 2 3 4), and (1 2)(3 4) having orders 1, 6, 8, 6, 3, respectively. Now, C_4 and C_5 contain no elements of S_3, while C_1, C_2, and C_3 contain the conjugacy classes T_1, T_2, and T_3 in S_3 of 1, (1 2), and (1 2 3). Note that $|C_i| = |T_i|$, $|C_2| = 2|T_2|$, $|C_3| = 4|T_3|$, and $[G : H] = 4$. Let ψ_1, ψ_2, and ψ_3 be the trivial, sign, and irreducible 2-dimensional characters of S_3, respectively. By Proposition 13.6.20, we obtain the following table from the character table of S_3:

S_4	e	(1 2)	(1 2 3)	(1 2 3 4)	(1 2)(3 4)
ϕ_1	4	2	1	0	0
ϕ_2	4	-2	1	0	0
ϕ_3	8	0	-1	0	0.

We will use this to determine the characters of S_4 once again. Assume we have already found its abelian characters, the trivial character χ_1 and the alternating character χ_2. Since

$$\langle \operatorname{Ind}_H^G \phi_i, \operatorname{Ind}_H^G \phi_i \rangle = \dim \operatorname{Hom}_{\mathbb{C}[G]}(V_i, V_i),$$

where V_i is the G-representation induced by ϕ_i, and these values are 2, 2, 3, respectively, we have that the V_i break up into these respective numbers of irreducible representations. But note that $\langle \phi_1, \chi_1 \rangle = 1$ and $\langle \phi_1, \chi_2 \rangle = 0$, so $\phi_1 - \chi_1$ so is an irreducible degree 3 character of G, which we previously called χ_4. Similarly, $\phi_2 - \chi_2$ is an irreducible degree 3 character, which we called χ_5. We compute that $\langle \phi_3, \chi_4 \rangle = \langle \phi_3, \chi_5 \rangle = 1$, and $\phi_3 - \chi_4 - \chi_5$ is an irreducible character of degree 2, which we called χ_3.

13.7. Applications to group theory

Let G be a group of order n, and let $C_1, \dots, C_r$ be the conjugacy classes in G, choose $g_j \in C_j$ and set $c_j = |C_j|$ for each $1 \le j \le r$. Let $\chi_1, \dots, \chi_r$ be the irreducible complex characters of G, and set $n_i = \deg \chi_i$ for $1 \le i \le r$. Let V_i denote the irreducible representation with character χ_i, let $\rho_i \colon \mathbb{C}[G] \to \operatorname{End}_{\mathbb{C}}(V_i)$ be the $\mathbb{C}$-algebra homomorphism restricting to the representation ρ_{V_i}. We also use χ_i to denote its $\mathbb{C}$-linear extension to map $\chi_i \colon \mathbb{C}[G] \to \mathbb{C}$.

PROPOSITION 13.7.1. *Set*

$$N_i = \{g \in G \mid \chi_i(g) = \chi_i(1)\}$$

for $1 \le i \le r$. *The normal subgroups of G are exactly the intersections* $\bigcap_{j \in J} N_j$, *where J is a subset of* $\{1, \dots, r\}$.

PROOF. First, N_i is a normal subgroup, since $\chi_i(g) = \chi_i(1)$ if and only if g acts as the identity, recalling that the eigenvalues of g are all roots of unity, so $N_i = \ker \rho_{V_i}$. It follows that every intersection of the N_i's is normal.

Now suppose N is normal in G. Let $V = \mathbb{C}[G/N]$. Let χ_V be the character of V as a $\mathbb{C}[G]$-module. Since $\ker \rho_V = N$, we have $\chi_V(g) = \chi_V(1)$ if and only if $g \in N$. If $g \notin N$, then $ghN \neq hN$ for any $h \in G$, so $\chi_V(g) = 0$.

Now χ_V is a sum of irreducible characters with nonnegative integer coefficients, say $\chi_V = \sum_{i=1}^{r} a_i \chi_i$. We claim that $N = \bigcap_{i \in J} N_i$, where J is the set of i with $a_i \geq 1$. Note that for any character ψ and $g \in G$, we have $|\psi(g)| \leq \psi(1)$ since $\psi(g)$ is a sum of $\psi(1)$ roots of unity. For $g \in G$, we have

$$\chi_V(g) = |\chi_V(g)| = \left| \sum_{i=1}^{r} a_i \chi_i(g) \right| \leq \sum_{i=1}^{r} a_i \chi_i(1) = \chi_V(1),$$

with equality of the first and last term holding if and only if $g \in N$. However, the middle inequality is an equality if and only if all $\chi_i(g)$ for $i \in J$ are equal and have absolute value $\chi_i(1)$. This condition holds if and only if all $\chi_i(g) = \chi_i(1)$, since one of these characters is the trivial character. □

Next, we will show how to find the center of G.

PROPOSITION 13.7.2. *Set*

$$Z_i = \{g \in G \mid |\chi_i(g)| = \chi_i(1)\}$$

for $1 \leq i \leq r$. *Then* Z_i *is a normal subgroup of* G, *and the center of* G *is equal to* $\bigcap_{i=1}^{r} Z_i$.

PROOF. We have that Z_i is a normal subgroup since the condition that $g \in Z_i$ is exactly that g acts as a scalar multiple of the identity, in other words that g is in the inverse image of the center of $\rho_{V_i}(G)$.

We claim that Z_i/N_i is the center of G/N_i. Note that ρ_{V_i} has kernel N_i defined as in Proposition 13.7.1, and the elements of Z_i are mapped to scalar matrices in the center of $\rho_{V_i}(G) \cong G/N_i$. So Z_i/N_i is contained in the center of G/N_i. Now suppose that gN_i is in the center of G/N_i. Then $\rho_{V_i}(g)$ commutes with all $\rho_{V_i}(h)$ for $h \in G$, which implies that left multiplication by g is a $\mathbb{C}[G]$-module isomorphism of V_i. But V_i is simple, so $\mathrm{Hom}_{\mathbb{C}[G]}(V_i, V_i) \cong \mathbb{C}$. In other words, g acts as scalar multiplication by some element, hence is contained in Z_i.

Given the claim, we have that $Z(G)N_i/N_i \leq Z_i/N_i$, and so $Z(G) \subseteq Z_i$ for all i. Now suppose that $z \in Z_i$ for all i. Let $g \in G$. Then $gzg^{-1}z^{-1} \in N_i$ by our earlier claim. But $\bigcap_{i=1}^{r} N_i$ is trivial by Proposition 13.7.1. So, $gzg^{-1}z^{-1} = 1$, for all $g \in G$, so $z \in Z(G)$, as desired. □

PROPOSITION 13.7.3. *For each pair* (i, j) *of integers with* $1 \leq i, j \leq r$ *and each* $g \in G$, *we have*

$$\frac{c_j}{n_i} \chi_i(g_j) \in \mathbb{Z}[\mu_n].$$

PROOF. Set $N_j = \sum_{g \in C_j} g \in Z(\mathbb{C}[G])$. Multiplication by N_j defines a $\mathbb{C}[G]$-linear map $N_j \colon V_i \to V_i$ which by Schur's lemma is a scalar multiple of the identity, say $N_j = \alpha_j \, \mathrm{id}_{V_i}$ for $\alpha_j \in \mathbb{C}$, which tells us that $\chi_i(N_j) = \alpha_j \chi_i(1) = \alpha_j n_i$. But we also have

$$\chi_i(N_j) = \sum_{g \in C_j} \chi_i(g) = c_j \chi_i(g_j),$$

so $\alpha_j = \frac{c_j}{n_i}\chi_i(g_j)$. We show that eac

Next, set

$$a_{jkl} = |\{(g,h) \mid g \in C_j, h \in C_k, gh = g_l\}| \in \mathbb{Z}_{\geq 0},$$

and note that this number is independent of the choice of g_l, since $sgs^{-1} \cdot shs^{-1} = sghs^{-1}$ for any $s \in G$. Then

$$\rho_i(N_j)\rho_i(N_k) = \sum_{g \in C_j} \sum_{h \in C_k} \rho_i(gh) = \sum_{l=1}^{r} a_{jkl} \sum_{q \in C_l} \rho_i(q) = \sum_{l=1}^{r} a_{jkl} \rho_i(N_l).$$

Since N_j acts on V_i by multiplication by the scalar α_j, this implies

$$\alpha_j \alpha_k = \sum_{l=1}^{r} a_{jkl} \alpha_l.$$

In particular the subring $\mathbb{Z}[\{\alpha_j \mid 1 \leq j \leq k\}]$ of $\mathbb{C}$ has finite $\mathbb{Z}$-rank, so it is integral over $\mathbb{Z}$. In particular, each α_j is integral over $\mathbb{Z}$. The result now follows as $\alpha_j \in \mathbb{Q}(\mu_n)$ for each j, and $\mathbb{Z}[\mu_n]$ is the integer ring of $\mathbb{Q}(\mu_n)$. □

COROLLARY 13.7.4. *The dimension of an irreducible complex representation of a finite group G divides $|G|$.*

PROOF. Let $1 \leq i \leq r$, and consider the quotient of interest

$$\frac{n}{n_i} = \frac{n}{n_i}\langle \chi_i, \chi_i \rangle = \sum_{j=1}^{r} \frac{c_j}{n_i} \chi_i(g_j)\overline{\chi_i(g_j)},$$

which is a $\mathbb{Z}[\mu_n]$-linear combination of the algebraic integers $\frac{c_j}{n_i}\chi_i(g_j)$, hence an algebraic integer. Since the fraction also lies in $\mathbb{Q}$, we have that n_i divides n. □